8-2016

Identification of Biomarkers Associated with Rous Sarcoma Virus-induced Tumors in Two Divergently Selected Chicken Lines

Ashley Hayden
University of Arkansas, Fayetteville

Identification of Biomarkers Associated with
Rous Sarcoma Virus-induced Tumors in Two Divergently Selected Chicken Lines

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Poultry Science

by

Ashley Hayden
University of Arkansas
Bachelor of Arts in Biology, 2011

August 2016
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Dr. Byung-Whi Kong
Thesis Director

_____
Dr. Nicholas Anthony
Thesis Co-Director

_____
Dr. Gisela Erf
Committee Member

_____
Dr. Douglas Rhoads
Committee Member

ABSTRACT

Poultry has become especially important to genetic research due to breeding feasibility, short generation turnover, and ease of maintaining large populations. The discovery of virus induced cancer has paved the way for further genetic studies. Rous Sarcoma Virus (RSV) is a tumor-causing virus that infects poultry. While not prevalent today, it can serve as a model for virus-induced cancer in humans and create additional insight to marker assisted selection in poultry. Genetically selected Arkansas Progressor (AP) and Arkansas Regressor (AR) chicken lines have been established and maintained at the Arkansas Experimental Station (AES) in Fayetteville, AR. Previous research has investigated the immunological and genetic characteristics of virus induced tumors. Publication of the Red Jungle Fowl genome has made it possible to perform genome wide studies to identify biomarkers associated with the susceptibility to RSV induced tumors. In this study, whole genomes of AP and AR lines were sequenced and analyzed using Illumina platform next generation sequencing. Over 9,000,000 SNPs were identified against the reference genome and 12,000 were classified as unique SNPs between the AR and AP line. Most SNPs were found in intergenic regions, while few were in protein coding regions. Unique SNPs were categorized by the following mutations: Frameshift (134), No Start (25), No Stop (7), Non-synonymous (6884), and Nonsense (112). Unique SNPs were characterized as occurring in only one line. Nine SNPs were chosen after genomic analyses for testing using PCR and Sanger sequencing in larger populations of AR and AP birds followed by validation in unrelated populations. No statistical correlations were found between the 9 SNPs in the unrelated populations, Giant Jungle Fowl (GJF) and White Leghorn (WL). No genetic difference was detected between birds of differing phenotypes within the GJF and WL lines. SNP frequencies were the same for both birds that regressed tumors and those that progressed tumors. Interestingly, it was observed that the GJF line response to RSV resembled the AR line and the WL line resembled the AP line, and GJF and WL exhibited differing phenotypes, indicating the potential to find biomarkers in larger population sizes.

ACKNOWLEDGEMENTS

I would like to thank both of my graduate advisors, Dr. Byung-Whi Kong and Dr. Nicholas Anthony. They have provided much guidance and encouragement throughout the entirety through this journey. They have both pushed me to work harder, whether it was trying to assemble a genome or taking tumor scores 3 times a week no matter the circumstances. Many thanks to Dr. Anthony for allowing me to use the Giant Jungle fowl, Arkansas Regressor, Arkansas Progressor lines and helping me find a White Leghorn population. Long days in the lab trying to get the perfect PCR results and those hot days in the chicken houses have paid off, not only for the purpose of this degree, but also for the work ethic and life experiences that shall remain with me.

I would also like to thank my committee members Dr. Douglas Rhoads and Dr. Gisela Erf. Dr. Rhoads helped me analyze SNP data, which was the starting point of my project. Without his help, I would have been sifting through data for weeks. Dr. Erf provided me with the knowledge of immunology, or more specifically, the mechanisms by which the immune system responds to viruses. This knowledge was crucial to this project. She also gave me advice on the proper handling and administration techniques for the virus used in this project.

Further, I would like to extend my thanks to my colleagues Audrianna, Alex, Grant, Katy and Kaylee, who gave up time to help me collect data three times a week for several trials. The farm crew (Lori Silva, Cory Burbridge and various other undergraduate students) also gave up some of their time to help collect samples and care for the birds throughout my trials. Bhuwan Khatri helped tremendously by helping me obtain PCR and sequencing results.

Lastly, I would like to thank my family who has been with me the entire way and has helped push and encourage me to complete this thesis. I would not have been able to accomplish this great achievement without them. Their support carried me through the tough times that arose during this process.

Most importantly, I would like to give thanks to God for allowing me the opportunity to grow in such an experience as this.

DEDICATION

I would like to dedicate this thesis to my sister and my brother-in-law, Jennah and Brian Bennett. They have always been there for me and helped pave the way toward many of my achievements in life. Without them, I would not be where I am today. I am truly thankful for all the love, guidance, and support they have given me for over thirteen years.

TABLE OF CONTENTS

INTRODUCTION

Identifying and understanding the genetic mechanisms by which evolutionary changes occur can provide insight into the association between the genotypic and phenotypic characteristics of organisms. Domestic animals serve as excellent models for genetic studies involving phenotypic evolution. Genetic adaptations have resulted in domestic animals' evolution under the control of human-driven selection leading to remarkable phenotypic changes. With the current advancements in molecular genetics, the identification of genetic markers has provided opportunities to improve conventional selection when the desired phenotype is not measurable and/or has low heritability (Dekkers, 2003). The studies described in this thesis will provide beneficial information toward marker-assisted selection.

This thesis will cover, in detail, the genetic basis of the chicken genome responsible for the regression and progression of tumors caused by RSV. The objective of this study was to analyze whole genome sequences of genetically selected chicken lines, Arkansas Progressor (AP) and Arkansas Regressor (AR), retaining commercially and biomedically important traits using next generation sequencing techniques and bioinformatics tools. Millions of SNPs were identified by genome sequencing for AP and AR lines and nine candidate genetic markers were tested for further application in unrelated populations. The main objective of this study was to find direct application of chosen SNPs for both genetic marker assisted animal selection and diagnosis of human genetic disorders.

CHAPTER 1: LITERATURE REVIEW

INTRODUCTION

Next generation sequencing has provided a stepping stone for studies in evolution, genetic diseases, and medicine. This thesis will demonstrate how the chicken can be used as a model for genetic investigations. Whole genome sequencing was used to identify potential genetic loci under selection for tumor progression and regression caused by Rous sarcoma virus (RSV) in chickens. RSV is a cancer-causing virus which induces malignant tumors of connective tissue known as sarcomas. Phenotypic data was collected and analyzed against genotypic data to identify potential markers associated with malignant sarcomas. In order to collect phenotypic data for the purposes of this study, the chicken was the chosen model.

BIOMARKER RESEARCH AND APPLICATION

Whole genome re-sequencing using next-generation massively parallel sequencing technologies greatly contributes to molecular applications in studying genetic variation. Current DNA sequencing technologies have improved throughput and reduced costs. Genome wide association studies provide more accurate results in mapping associated loci when compared to the alternative technique, linkage analysis. One such study identified a region of the chicken genome containing the genes ROBO1 and ROBO2, which were found to strongly affect antibody response to Newcastle disease in chickens (Luo et al., 2013). Genetic polymorphisms provide evidence of variations in specific phenotypes such as disease resistance, response to drugs or the environment, or improved performance (Li et al., 2009). Dekkers et al. (2007) found that using marker-assisted selection in commercial crossbred chickens increased performance by 10% and decreased inbreeding from 2.1% to 1.4%. The use of these technologies allows the discovery of genetic mechanisms that can lead to phenotypic differentiation by identifying genomic mutations in wide ranges of genomic regions, which have been accumulated under selection (Johansson et al., 2010).

Whole genome sequencing has also greatly benefited biomedical science in humans. This advancement enabled scientists to discover biomarkers, which can aid in the diagnosis and prognosis of certain diseases. Two genes, miRNA-21 and KLK, were discovered to be potential biomarkers associated

with the response in cancer immunotherapy (Wang, 2013) and prostate cancer prognosis (Batra et al., 2012), respectively. Today, single nucleotide polymorphisms (SNPs) are being used as candidate biomarkers which create alternative forms of genotypes. A population containing different genotypes for one particular loci could create distinct phenotypes causing variation within the population. With technology continuing to advance, the association between genotypes and their causal phenotypes will contribute to marker assisted genetic selection and disease diagnosis and prevention.

CHICKEN AS A MODEL FOR GENETIC RESEARCH

The advancement of knowledge in the etiology and pathogenesis of human diseases, including cancer, heart diseases and various infectious diseases, have been futhered by the use of animal models. The chicken has a long history in oncology studies, especially those associated with oncogenic viruses (Calnek, 1992). In the early 1900s, cell-free filtrates were used to transmit leukemias and sarcomas in chickens (Rous, 1911). More recently, studies with avian luekosis/sarcoma viruses have paved the road in the first attempts to identify the genetic and molecular basis of viral oncogenesis (Neel et al, 1981).

Research by Rubin (1960) led to the discovery of the resistance-inducing factor (RIF) test for avian leukosis virus (ALV), which later allowed definition of the epidemiology, pathogenesis and immunology of lymphoid leukosis (Cooper et al., 1968). From this study, target cells and characteristics of transformed cells and viruses were investigated in addition to the host-virus interactions. RSV was the first virus identified to be transmissible in chickens (Weiss et al., 2011). Hanafusa et al. (1964) was able to discover the ‘src’ oncogene in the chicken genome after Rubin’s finding of RIF. The ‘src’ gene produces a protein which is a 60,000 dalton-phosphoprotein ($pp60^{src}$). This protein is necessary for transforming cells to sarcomas in birds. Sefton et al. (1979) inferred that $pp60^{src}$ is a protein kinase and the phosphorylation of tyrosine in $pp60^{src}$ is vital to the RSV induced malignant transformation of cells. The $pp60^{src}$ protein was phosphorylated in the tyrosine residues in both human and chicken cells when examined by immunoprecipitation assay. Moreover, all uninfected vertebrate cells contained $pp60^{src}$, a homolog of viral $pp60^{src}$ in the host genome, termed as a cellular src (cSRC) and cSRC was characterized as a

cellular proto-oncogene (Sefton et al., 1986). The prevalence and variety of naturally-occurring sarcomas in chickens contribute to the validity in using the chicken as a model for tumor and cancer research.

RSV BACKGROUND AND MECHANISM

RSV is in the family *Retroviridae* and subfamily *Orthoretrovirinae* which consists of all oncogenic retroviruses. Oncogenic viruses are further divided into seven genera based on morphology and genome complexity. The seven genera are divided into simple and complex retroviruses. The simple retroviruses include *alpharetroviruses, betaretroviruses* and *gammaretroviruses*. RSV is an example of a retrovirus in the genera *alpharetrovirus*. Simple retrovirus genomes encode three genes: virion capsid/nucleopcapsid (Gag) proteins responsible for core proteins, Pol genes, which are enzymes to aid in genome replication (reverse transcriptase and integrase), and envelope (Env) proteins to enable cell infection (Braoudaki et al., 2011). Reverse transcriptase becomes active upon penetration of the host cell and the virion uncoats. Primers hybridized to the viral RNA genome are elongated to form a complimentary DNA, and then the reverse transcriptase removes the parental RNA from the DNA. A second strand of DNA is synthesized to generate a double-stranded DNA provirus. The provirus integrates itself into host chromosomal DNA catalyzed by the viral Integrase. When the host cell proliferates, the provirus is replicated along with the host DNA (Weiss, 1996). If the provirus genome is expressed (transcribed), that cell will transform into a tumor cell and can proliferate and form a sarcoma.

Sarcomas are solid tumors derived from bone, cartilage, fat, muscle, vascular, or hematopoietic tissues that can metastasize in any organ of the body (Braoudaki et al., 2011). The study described in this thesis uses RSV to induce sarcomas in the wing web of a chicken to study genotypic-phenotypic relationships in the chicken genome to find possible mutations associated with RSV response. Mutations in host oncogenes can be the cause for the susceptibility to RSV-induced tumors.

CELLULAR RESPONSE TO VIRUSES

A host cell initially detects a virus using the innate immune system. The innate immune system is responsible for the first 5 days of immunological activity against a virus. Viruses, such as RSV typically infect cells through attachment to normal cell surface molecules as receptors, thereby enabling entry into the innate cell. Receptors of innate cells recognize unique molecular structures expressed by different classes of infectious organisms. Cells of the innate immunesystem include monocytic cells, dendritic cells (DC), natural killer (NK) cells, and polymorphonuclear leukocytes (PMN). The innate immune system mediates a response using preexisting cells to direct and support other protective responses (Shizuo et al. 2006).

The key mechanism s of the innate system against viruses are inhibiting infection by type I interferons and NK cell-mediated killing of the infected cells. The production of type I interferons is triggered by several biochemical pathways. Recognition of viral RNA and DNA by endosomal Toll-like receptors (TLRs) and activation of cytoplasmic RIG-like receptors by viral RNA is followed by the convergence of these two pathways on the activation of protein kinases. The protein kinases activate the IRF (Interferon regulatory factors) transcription factors which in turn stimulate interferon gene transcription whose products interfere with viral replication. NK cells kill infected cells before the adaptive immune responses develop. NK cells recognize infected cells in which a virus has turned off class I MHC expression as a mechanism to escape from cytotoxic T cells (CTLs). The absence of class I MHC releases NK cells from a normal state of inhibition. The most common CTLs involved in viral infection are $CD8^+$ T cells which recognize cytosolic viral peptides presented by class I MHC molecules. The primary role of CTLs is to kill any infected nucleated cells. The antiviral activity of CTLs comes from killing the cells but also activating nucleases within infected cells. These nucleases can degrade viral genomes and secretion of cytokines which can activate phagocytes (Perry et al, 2005). Identification of genetic mutations housed in genes which produce proteins involved in this process can increase knowledge of the immune system, in turn contributing to possible advancements in medicine.

HISTORY OF LINEAGE FOR AR AND AP LINES

The two divergently selected chicken lines used in this study are the Arkansas Regressor (AR) and Arkansas Progressor (AP) lines. The AR and AP lines were developed from purebred White Leghorn (WL) and Giant Jungle Fowl (GJF) strains (Gyles et al., 1967). The WL strain was maintained at the Arkansas Agricultural Experiment Station (AES) in Fayetteville, AR as a small flock, beginning in 1946. Due to the parental population size, this line is highly inbred but the inbreeding coefficient is unknown. The GJF strain's origin stems from one male and five hens and is theorized to be highly inbred also. The GJF male and five hens were transported to Fayetteville, AR from Southeast Asia in 1951 (Gyles et al., 1966). GJF line has been maintained at the AES since 1951 as a small closed flock.

In 1964, the chorioallantoic membrane (CAM) of chick embryo was inoculated with RSV of both the WL and GJF strains. WL was discovered to be highly susceptible to tumor development by inoculations of RSV (Bower, 1962; Bower et al., 1964) while GJF showed high resistance to tumor development to inoculations of RSV. These findings led to the development of the AP and AR lines used in this thesis (Gyles et al., 1967; Bower, 1962; Bower, 1962; Bower et al., 1964). The GJF and the WL strains were crossed in 1965 to produce the AR line maintained at the AES in Fayetteville, AR. One reciprocal F_1 cross (GJF X WL) was created by the mating of 8 GJF males to 10 WL hens. The other reciprocal F_1 cross (WL X GJF) was created by the mating of 9 WL males to 17 GJF hens. The AP line consists of only WL paired matings and was created with 13 hens to 10 males (Gyles et al. 1966). All matings were pedigreed and performed by artificial insemination.

GENETIC RESISTANCE TO RSV: MHC AND NON-MHC GENES

There is a strong association between the chicken Major Histocompatibility Complex (MHC) and disease caused by small infectious pathogens (Collins et al., 1977). The chicken MHC is composed of single dominantly expressed genes resulting in decreased peptide recognition by immune receptors. In the case of genetic resistance to RSV to chickens, the B-F/B-L region of the B locus is the major determinant of regression and progression of RSV-induced tumors (Kaufman et al., 2000). Gyles et al.

(1977) tested this genetic association to MHC by mating chickens selectively based on tumor progression or tumor regression. The progressor mating line was purebred White Leghorn while the regressor line was comprised of White Leghorn and Giant Jungle Fowl (GJF) and their crosses. This study suggested that there was great genetic influence on tumor growth caused by RSV.

Several studies have shown that the B complex in chickens is comparable to the MHC molecules in other species, including humans. However, the mammalian MHC has greater genetic variation as a result of more balanced selection (Aquilar et al., 2003). The chicken B complex contains 19 genes compacted into a 92-kb region, making it much smaller than the Human Leukocyte Antigen (HLA), the human MHC. The small size and simplicity of the chicken MHC makes it useful in studying pathogen-specific disease resistance (Kaufman, 1999). Kaufman et al. (1999) has sequenced the chicken MHC complex and has revealed that there are minimal counterpart genes in the HLA, signifying conservation of the chicken MHC for over 200 million years, during which time the divergence of mammals and birds occurred.

The MHC complex consists of genes which code for cell surface receptors responsible for adaptive immune response. Class I MHC molecules are present on all nucleated cells and class II MHC molecules are present only on antigen presenting cells (APC's). The most common APC's include dendritic cells, macrophages and B lymphocytes. These cells express the MHC II complex which displays a peptide that binds to the T-cell Receptor (TCR) of CD4+ T lymphocytes, also known as Helper T-cells. The TCR recognizes the peptides via amino acid sequence; The TCR's are specific for one peptide displayed on one antigen. Once a TCR recognizes a peptide, the T-cell is activated leading to clonal expansion and differentiation, activation of macrophages, and activation of B-cells for antibody production. The polymorphism of the MHC genes creates the specificity and diversity of the antigen recognition process performed by T-cells which stimulates the immune response (Abbas, 2012). It is important to have a high number of specificities in one individual's MHC genes to ensure healthy immunity against pathogens. Class I MHC molecules and an endogenous antigen peptide are recognized by the TCR of CD8+ T lymphocytes to become activated cytotoxic T lymphocytes (CTLs). CTLs are specialized to kill cells infected with intracellular microbes, including viruses, and tumors expressing tumor antigens (Abbas, 2012). When infected with RSV, the genetic variation, or lack thereof, in class I

molecules influence the cellular response. If there is not a strong binding between the TCR of the CD8+ T-cell and the antigen being presented by the class I MHC molecule, infected cells can proliferate and form sarcomas. The genetic variation in the MHC region is crucial to infection and disease resistance.

For an immune response to occur, antigen-specific receptors must recognize antigens and illicit a response between T-cells and macrophages, dendritic cell, or B-cells with one product being antibodies. As previously explained, MHC molecules play an important role in immune response to pathogens, and are restricted to certain identities making the cells specific for one antigen such as a viral subgroup of Rous Sarcoma Virus (Taylor, 2004). There are two classes of chicken MHC molecules and they show to be analogous to those of mammals: the B-F and B-L classes of MHC molecules, making the chicken a great model for immunological studies. However, the chicken has a third unique class of MHC molecule named the B-G class. Additional genes in the chicken MHC region involved in lymphocyte and TAP2 (transporter associated with antigen procession) activation are also homologous to most mammalian genes. In the chicken, the MHC antigens are also required for cellular communication in immune responses. The peptides that are presented by the MHC molecules are different for each individual with genetically different MHC sequences. This feature allows a broader spectrum of immune responses against diseases as well as maintains the individual from self-destruction (Lamont, 1998).

Today, many agree that not one single mechanism is accountable for these large differences among individuals. Selection and recombination together fulfill the role of MHC generation (Hess et al. 2002). The selection involved in the creation of the MHC polymorphisms of these genes is balanced by two different types of selection: overdominant selection and frequency-dependent selection (Parham et al. 1996). Overdominant selection is the process by which individuals that are heterozygous at certain loci are more likely to survive and have higher fitness. This is true for MHC genes because if there are two different haplotypes expressed for a pathogen, that individual can elicit a more effective immune response against a larger number of pathogens than an individual homozygous at that locus. Frequency-dependent selection results when a certain allele in a population is strongly selected due to an epidemic, in turn causing a more effective immune response against a particular pathogen. After time, this allele may increase in frequency until it dominates the locus, known as fixation (Parham et al. 1996). In this study, it was hypothesized that the AP and AR chicken lines would show several fixations in the genome

due to a similar scenario. In the areas of fixation, the two lines would have differing alleles, as they were selected divergently based on phenotype.

MHC genes associated with resistance to RSV induced tumor developments

There are two sub regions of the chicken MHC complex, the B and Y complexes, assort independently, each contain MHC class I and MHC class II genes, separated by the nucleolar organizer region (NOR) (Miller et al., 1996). Hosomichi et al. (2008) found that the B-complex of chickens, just as the human HLA, derives its diversity using several different mutational events. Hosomichi et al. (2008) sequenced a 14-gene region in the chicken MHC, which was a total of 59kb ranging from BG1 to BF2 on the genome. The results suggested that some of the haplotypes were created from whole and partial allelic gene conversion as well as homologous reciprocal recombination on top of the nucleotide mutations. These haplotypes associate with some diseases that compromise the immune system; however it is still unknown as to specifically which genes are associated to which disease responses (Hosomichi et al., 2008).

There has been little research investigating the role of the Y complex in RSV resistance. However, it has been determined that different Y genotypes influence tumor growth pattern, fate of tumor, and mortality rates in chickens having a B^2B^5 background and the three different *Y* alleles: $Y^{1.1}$, $Y^{1.2}$, *and* Y^6 (LePage et al., 2000). Another study performed by Pinard-Van der Laan et al. (2004) also indicated that the Y complex had a small role in RSV resistance in two chicken lines selected for tumor progression or regression (not the AR and AP lines used for this study). An overall tumor profile index (TPI) heritability of 0.46±0.03 was observed for both regressor and progressor lines having the same Y genotype. The study found a significant influence of Y genotype only in the progressor line. However, it was later discovered that the progressor line from Pinar-Van der Laan's study had the same sensitivity in the Y-genotype as the regressor line. Praharaj et. al (2004) concluded that the mechanism of the response was different between the two lines. The progressor line exhibited an anti-tumoral response while the regressor line exhibited more of an antiviral response. This led scientists to believe that non-MHC genes are associated with tumor development caused by RSV. This study aims to identify any genes, MHC or non-MHC that can be correlated to RSV-induced tumor progression and regression.

Non-MHC genes associated with resistance to RSV induced tumor development

Collins et al. (1985) discovered the involvement of non-MHC genes in the immune response of RSV. The study monitored the tumor incidence and metastasis in chickens with several different haplotypes. Some chickens with the same haplotypes had completely different rates of metastasis, for example: 52 B^5/B^5 individuals had an average metastasis rate of 60% while another group of 120 individuals with the same genotype had a rate of 31%. This indicated that a non-MHC genetic effect was taking place on the metastasis of those chickens (Collins et al., 1985).

The specific loci involved in the association of RSV with the non-MHC genes are still being studied. Gilmour et al. (1983) examined the effects of two non-MHC alloantigens, *Ly-4* and *Th-1*, on chicken lines (6_3 x 7_2) F_5 birds (B^2B^2). The *Ly-4*^{a}*Ly-4*^{a}/*Th-1*^{a}*Th-1*^{a} genotype showed significantly lower tumor incidence. High tumor incidence resulted in the presence of *Ly-4*^{a} and *Bu-1*^{b} alleles and also the *Ly-4*^{a} and *Bu-1*^{a} alleles in line (6_3 x 100) F_4 progeny. This line had *Ly-4* and *Bu-1* segregated on a shared B^2B^2 background (Gilmour et al. 1986). This thesis will further discuss the possible non-MHC genes involved in RSV induced metastasis in chickens.

MARKERS ASSOCIATED WITH TUMORS AND CANCER

Technological advances such as PCR, next generation sequencing, SNP chips, and gene expression assays have decreased the cost and efforts in biomarker discovery. These techniques have helped identify biomarkers that aid in cancer prognoses, diagnoses, and treatments. For many cancers, tumors develop due to genetic mutations affecting protein synthesis. The human papilloma virus (HPV) causes cancer and/or growths, similar to RSV. Based on data from 2006-2010, it is estimated that 33,200 HPV-related cancers occur each year in the United States. Recent studies show that HPV is responsible for more than 90% of cervical cancer and is associated with roughly 70% of cancers of the oropharynx (Division of Cancer Prevention and Control, Centers for Disease Control and Prevention, 2014). Two genes, C4.8 and C21.7, were found to be present in over 58% of all cancers, being potential biomarkers in early cancer prognosis (Nees et al., 1998). In 2012, Next Generation Sequencing (NGS) enabled discovery of 7241 somatic point mutations, mutations in 40 cancer genes and 73 combinations of mutated

cancer genes (Stephens et al., 2012). Studying the genetic variations contributing to RSV-induced tumor growth can add valuable knowledge toward cancer research. Biomarkers can help scientists pinpoint areas of the human genome responsible for the onset of cancer, thus getting close to finding a cure for cancer.

NEXT GENERATION SEQUENCING AND SNP DISCOVERY

Next generation sequencing (NGS) has made whole-genome sequencing attainable and practical for researchers. Advantages of NGS include: higher resolution, more accurate readings, and large volumes of data in a short amount of time. One method of NGS, whole genome sequencing, is especially ideal for identifying causative variants providing scalability and flexibility (Illumina, 2015).

Next generation sequencing is commonly applied to livestock populations to increase economic value. Today geneticists select agricultural animals to improve performance and lower disease risk. Single Nucleotide Polymorphisms (SNPs) have become particular interesting in studying genotypic-phenotypic relationships. One SNP or a combination of SNPs can be indicative for a phenotype, such as growth rate, egg production, or disease susceptibility. If a SNP is found to be fixed and is associated with a particular phenotype, geneticists can use that SNP as a biomarker to improve selective breeding in chickens. Molecular genetics and traditional genetics are both being used together in selection techniques in breeding programs and will continue complementing each other while technological advances continue (Emmerson, 1997).

In 2004, the first draft of the *Gallus gallus* genome was published (v2.1). The most recent version (v. 4.0) was released in 2011 (International Chicken Genome Sequencing Consortium, 2004). The genome assembly is for one female Red Jungle Fowl (RFJ) from the UCD001 line and is considered the closest living ancestor of the domestic chicken. The reference genome is approximately 1,000 Mb in length and contains approximately 20,000 genes across 38 autosomes and one pair of sex chromosomes. The most recent assembly of *Homo sapiens* is approximately 3,200Mb in length and contains 20,000-25,000 protein-coding genes (International Chicken Genome Sequencing Consortium, 2004), making the chicken genome one third its size. The publication of the whole genome of the chicken

has greatly influenced molecular genetics, providing major advances not only in the medical field, but agricultural research as well.

SUMMARY

In this thesis, next generation sequencing and annotation was performed for AR and AP lines using the RJF genome as a reference (International Chicken Genome Sequencing Consortium, 2004). Using this technique, SNPs were identified as candidate biomarkers associated with specific phenotypes of the AR and AP lines. The selection of SNP's was based on gene function and diverging SNP frequencies between the two lines. The SNP's were tested in two different unrelated populations, GJF and WL to determine whether or not the chosen SNPs were indeed potential biomarkers for predicting the fate of sarcomas in chickens. Results indicated the chosen SNPs were not indicative of the regression or progression of RSV-induced tumors. Although, SNPs were not validated within each line, the GJF and WL exhibited genetic differences in SNP frequencies and grand mean tumor scores in relation to each other.

REFERENCES

Abbas, Abul K., Andrew H. Lichtman, and S. Pillai. (2012) Cellular and Molecular Immunology. 7th ed. Philadelphia: Elsevier/Saunders, 110-111.

Afanassieff, M., R. M. Goto, J. Ha, M. A. Sherman, L. Zhong, C. Auffray, F. Coudert, R. Zoorob, and M. M. Miller. (2001). At Least One Class I Gene in Restriction Fragment Pattern-Y (Rfp-Y), the Second MHC Gene Cluster in the Chicken, Is Transcribed, Polymorphic, and Shows Divergent Specialization in Antigen Binding Region. *The Journal of Immunology*. 3324-333.

Aguilar, A., G. Roemer, S. Debenham, M. Binns, D. Garcelon, and R. K. Wayne. (2004). High MHC Diversity Maintained by Balancing Selection in an Otherwise Genetically Monomorphic Mammal. *Proceedings of the National Academy of Sciences* 101.10. 3490-494.

Bacon, L. D., H. D. Hunt, and H. H. Cheng. (2000). A Review of the Development of Chicken Lines to Resolve Genes Determining Resistance to Diseases*. Poultry Science* 1082-093.

Ball, L. A. Fields *Virology*. (2000) Ed. D. Knipe and P. Howley. 5th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 120-140.

Batra, Jyotsna, Tracy O'Mara, Radhika Patnala, Felicity Lose, and Judith A. Clements. (2012).Genetic Polymorphisms in the Human Tissue Kallikrein (KLK) Locus and Their Implication in Various Malignant and Non-malignant Diseases. *Biological Chemistry.*

Beard, J. W. (1958). Viruses as the cause of cancer. *American Scientist*. 226-254

Bloom, S. E., Bacon, L. D. (1985). Linkage of the major histocompatibility (B) complex and the nucleolar organizer region in the chicken. *Journal of Heredity*. 76:146-54.

Bower, Raymond K. (1962). A Quantitation of the Influence of the Chick Embryo Genotype on Tumor Production by Rous Sarcoma Virus on the Chorioallantoic Membrane. *Virology* 18: 372-77.

Bower, R. K., Gyles, N. R. and Brown, C. J. (1965). The number of genes controlling the response of chick embryo chorioallantoic membranes to tumor induction by Rous sarcoma virus. *Genetics* 51:739- 46.

Bower, Raymond K., N. Roy Gyles, and Connell J. Brown. (1964). Tumor Induction by Rous Sarcoma Virus on the Chorioallantoic Membranes of Reciprocal Crosses between Resistant and Susceptible Strains of Chickens. *Virology 47-50.*

Braoudaki, Maria, and Fotini Tzortzatou-Stathopoulou. (2011).Tumorigenesis Related to Retroviral Infections. *The Journal of Infection in Developing Countries*.

Calnek, B. W. (1992). Chicken Neoplasia—a Model for Cancer Research 1. *British Poultry Science*. 3-16.

Collins, Walter M., W. Elwood Briles, Robert M. Zsigray, William R. Dunlop, Allan C. Corbett, Kathy K. Clark, James L. Marks, and Timothy P. Mcgrail. (1977). The B Locus (MHC) in the Chicken: Association with the Fate of RSV-induced Tumors*. ImmunoGenetics*. 333-43.

Collins, Walter M., David W. Brown, Paul H. Ward, William R. Dunlop, and W. Elwood Briles. (1985). MHC and Non-MHC Genetic Influences on Rous Sarcoma Metastasis in Chickens*. ImmunoGenetics*. 315-21.

Dekkers, J. C. M. (2007). Marker-assisted Selection for Commercial Crossbred Performance. *Journal of Animal Science* 85.9. 2104-114.

Dekkers, J.C.M. (2004). Commercial Application of Marker- and Gene-assisted Selection in Livestock: Strategies and Lessons. *Journal of Animal Science* 82.13. E313-328.

Emmerson, D. (1997). Commercial Approaches to Genetic Selection for Growth and Feed Conversion in Domestic Poultry. *Poultry Science*. 1121-125.

Gilmour, Douglas G., Walter M. Collins, Tom L. Fredericksen, Willard E. Urban, Paul F. Ward, and Nancy L. Difronzo. (1986). Genetic Interaction between Non-MHC T- and B-cell Alloantigens in Response to Rous Sarcomas in Chickens. *ImmunoGenetics* 1-6.

Gilmour, Douglas G., Walter M. Collins, Tom L. Fredericksen, Billye Auclair, Kathy K. Clark, and Willard E. Urban. (1983). Influence of Non-MHC T Lymphocyte Alloantigens on Regression of Rous Sarcomas in the Chicken*. ImmunoGenetics* 43-54.

Gyles, N. R. (1966). A genetic model for the natural regression of tumors? *Poultry Science*. 30-56.

Gyles, N. R., M. Blythe, P. Test, A. Bowanko, and C. J. Brown. (1977). Immune Response in Progressor and Regressor Strains of Chickens at Specific Intervals After a Primary Challenge with Rous Sarcoma Virus. *Poultry Science* 758-66.

Gyles, N. R., J. L. Miley, and C. J. Brown. (1967). Size, Score and Speed of Development of Tumors in Chickens in Response to Subcutaneous Inoculations with Rous Sarcoma Virus. *Poultry Science*. 789-95.

Hanafusa, H., T. Hanafusa, and H. Rubin. (1964). Analysis Of The Defectiveness Of Rous Sarcoma Virus, Ii. Specification Of Rsv Antigenicity By Helper Virus. *Proceedings of the National Academy of Sciences.* 41-48.

Hess, Christopher M., and Scott V. Edwards. (2002). The Evolution of the Major Histocompatibility Complex in Birds*. BioScience* 423.

Hosomichi, K., M. M. Miller, R. M. Goto, Y. Wang, S. Suzuki, J. K. Kulski, M. Nishibori, H. Inoko, K. Hanzawa, and T. Shiina. (2008). Contribution of Mutation, Recombination, and Gene Conversion to Chicken Mhc-B Haplotype Diversity. *The Journal of Immunology*. 3393-399.

HPV-Associated Cancers Statistics. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, 02 Sept. 2014. Web. 15 Nov. 2015.

Johansson, Anna M., Mats E. Pettersson, Paul B. Siegel, and Örjan Carlborg. (2010). Genome-Wide Effects of Long-Term Divergent Selection*. PLoS Genetics*

Kaufman, J. (2000). The Simple Chicken Major Histocompatibility Complex: Life and Death in the Face of Pathogens and Vaccines. *Philosophical Transactions of the Royal Society: Biological Sciences*. 1077-084.

Kaufman, J., Miline, S., Gobel, T. W., Walker, B. A., Jacob, J. P., Auffray, C., Zoorob, R., and Beck, S. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature.* 401 (6756):923-925.

Lamont, S. J. (1998). Impact of Genetics on Disease Resistance. *Poultry Science*.1111-118.

Le Page, Keith T., Marcia M. Miller, W. Elwood Briles, and R. L. Taylor Jr. (2000). Rfp-Y Genotype Affects the Fate of Rous Sarcomas in B 2 B 5 Chickens. *ImmunoGenetics*. 751-54.

Li, R., Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, and J. Wang. (2009). SNP Detection for Massively Parallel Whole-genome Resequencing. *Genome Research*. 1124-132.

Luo, Chenglong, Hao Qu, Jie Ma, Jie Wang, Chunyu Li, Chunfen Yang, Xiaoxiang Hu, Ning Li, and Dingming Shu. (2013). Genome-wide Association Study of Antibody Response to Newcastle Disease Virus in Chicken. *BMC Genetics.* 14.1 42.

Miller, M. M., R. M. Goto, R. L. Taylor, R. Zoorob, C. Auffray, R. W. Briles, W. E. Briles, and S. E. Bloom. (1996). Assignment of Rfp-Y to the Chicken Major Histocompatibility Complex/NOR Microchromosome and Evidence for High-frequency Recombination Associated with the Nucleolar Organizer Region. *Proceedings of the National Academy of Sciences*. 93.9 3958-962.

Neel, Benjamin G., William S. Hayward, Harriet L. Robinson, Joanna Fang, and Susan M. Astrin. (1981). Avian Leukosis Virus-induced Tumors Have Common Proviral Integration Sites and Synthesize Discrete New RNAs: Oncogenesis by Promoter Insertion. *Cell* 23.2. 323-34.

Nees, Matthias, Erwin Van Wijngaarden, Evi Bakos, Achim Schneider, and Matthias Dürst. (1998). Identification of Novel Molecular Markers Which Correlate with HPV-induced Tumor Progression. *Oncogene* 16.19 2447-458.

Next-Generation Sequencing (NGS). Next-Generation Sequencing (NGS). Illumina, Inc., 2015. http://www.illumina.com/content/illumina-marketing/us/en/techniques/sequencing/dna-sequencing/whole-genome-sequencing.html Illumina 2015. Web. 15 Nov. 2015.

Parham, P., and T. Ohta. (1996). Population Biology of Antigen Presentation by MHC Class I Molecules. *Science* 67-74.

Pinard-Van der Laan, M. H., Soubieux, D., Merat, L., Bouret, D., Luneau, G., Dambrine, G., and Thoraval, P. (2004). Genetic analysis of a divergent selection for resistance to Rous sarcomas in chickens. *Genetics Selection Evolution.* 36(1): 65-81.

Praharaj, N., C. Beaumont, G. Dambrine, D. Soubieux, L. Merat, D. Bouret, G. Luneau, J. M. Alletru, M. H. Pinard-Van Der Laan, P. Thoraval, and S. Mignon-Grasteau. (2004). Genetic Analysis of the Growth Curve of Rous Sarcoma Virus-induced Tumors in Chickens. *Poultry Science*.1479-488.

Rous, P. (1911). A Sarcoma Of The Fowl Transmissible By An Agent Separable From The Tumor Cells. *Journal of Experimental Medicine*. 397-411.

Rubin, Harry. (1960). The Suppression of Morphological Alterations in Cells Infected with Rous Sarcoma Virus. *Virology*. 14-31.

Sefton, B. M., and T. Hunter. (1986). From C-src to V-src, or the Case of the Missing C Terminus*. Cancer Survey* 5.2. 159-72.

Sefton, B. M., Hunter, T. and Beemon, K. (1979.). Product of in vitro translation of the Rous sarcomus virus src gene has protein kinase activity. *Journal of Virology*. 30 (1): 311-318.

Akira, Shizuo, Satoshi Uematsu, and Osamu Takeuchi. (2006). Pathogen Recognition And Innate Immunity. *Cell.* 783-801.

Spanakos, Manolis V. (2007). The effects of MHC and non-MHC genes on Rous sarcoma virus tumor development in the Arkansas progressor and regressor chicken lines. *ProQuest.*

Stephens, Philip J., Patrick S. Tarpey, Helen Davies, Peter Van Loo, Chris Greenman, David C. Wedge, Serena Nik Zainal, Sancha Martin, Ignacio Varela, Graham R. Bignell, Lucy R. Yates, Elli Papaemmanuil, David Beare, Adam Butler, Angela Cheverton, John Gamble, Jonathan Hinton,

Mingming Jia, Alagu Jayakumar, David Jones, Calli Latimer, King Wai Lau, Stuart Mclaren, David J. Mcbride, Andrew Menzies, Laura Mudie, Keiran Raine, Roland Rad, Michael Spencer Chapman, Jon Teague, Douglas Easton, Anita Langerød, Osbreac, Ming Ta Michael Lee, Chen-Yang Shen, Benita Tan Kiat Tee, Bernice Wong Huimin, Annegien Broeks, Ana Cristina Vargas, Gulisa Turashvili, John Martens, Aquila Fatima, Penelope Miron, Suet-Feung Chin, Gilles Thomas, Sandrine Boyault, Odette Mariani, Sunil R. Lakhani, Marc Van De Vijver, Laura Van 'T Veer, John Foekens, Christine Desmedt, Christos Sotiriou, Andrew Tutt, Carlos Caldas, Jorge S. Reis-Filho, Samuel A. J. R. Aparicio, Anne Vincent Salomon, Anne-Lise Børresen-Dale, Andrea Richardson, Peter J. Campbell, P. Andrew Futreal, Michael R. Stratton, Rolf Karesen, Ellen Schlichting, Bjorn Naume, Torill Sauer, and Lars Ottestad. (2012). The Landscape of Cancer Genes and Mutational Processes in Breast Cancer. *Nature*. 400-404.

Strand, Tanja M., Gernot Segelbacher, María Quintela, Lingyun Xiao, Tomas Axelsson, and Jacob Höglund. (2012). Can Balancing Selection on MHC Loci Counteract Genetic Drift in Small Fragmented Populations of Black Grouse? *Ecology and Evolution*. 341-53.

Taylor, R. (2004). Major Histocompatibility (B) Complex Control of Responses against Rous Sarcomas. *Poultry Science*. 638-49.

Thoraval, Pierrick, Marielle Afanassieff, Daniele Bouret, Gillette Luneau, Evelyne Esnault, Ronald M. Goto, Anne-Marie Chausse, Rima Zoorob, Denis Soubieux, Marcia M. Miller, and Ginette Dambrine. (2003). Role of Nonclassical Class I Genes of the Chicken Major Histocompatibility Complex Rfp-Y Locus in Transplantation Immunity. *ImmunoGenetics.* 647-51.

Viruses That Can Lead to Cancer. Sept. 2014. Viruses That Can Lead to Cancer. American Cancer Association, 29 Web. 15 Nov. 2015.

Wang, Zhiyu, Jinxiang Han, Yazhou Cui, Kaixi Fan, and Xiaoyan Zhou. (2013). Circulating MicroRNA-21 as Noninvasive Predictive Biomarker for Response in Cancer Immunotherapy. *Medical Hypotheses*. 41-43.

Weiss, R. A. (1996). Retrovirus Classification and Cell Interactions. *Journal of Antimicrobial Chemotherapy* 37.Suppl B. 1-11.

Weiss, R. A., and P. K. Vogt. (2011). 100 Years of Rous Sarcoma Virus. *Journal of Experimental Medicine* 208.12. 2351-355.

Zhao, Mian, Yongzhen Wang, Hang Shen, Chenliang Li, Cheng Chen, Zhenhua Luo, and Hua Wu. (2013). Evolution by Selection, Recombination, and Gene Duplication in MHC Class I Genes of Two Rhacophoridae Species. *BMC Evolutionary Biology*. 113.

Zoorob, Rima, Alain Bernot, Dominique Martel Renoir, Fatima Choukri, and Charles Auffray. . (1993). Chicken Major Histocompatibility Complex Class II B Genes: Analysis of Interallelic and Inter-locus Sequence Variance. *European Journal of* Immunology. 1139-145.

CHAPTER 2: WHOLE GENOME SEQUENCING AND DATA ANALYSIS FOR AP AND AR LINES

INTRODUCTION

Whole genome sequencing is a useful method to study evolutionary mechanisms and genetic variation (Amaral et al, 2011). With technological advances in genetics and genomics, whole genomes can be sequenced at a relatively low cost and in a short amount of time. Whole genomes for several organisms, including the ancestral wild chicken, Red jungle fowl, have been successfully sequenced and published (International Chicken Genome Consortium). Massively parallel next-generation sequencing technologies for large-scale re-sequencing of cohorts of individuals with a known reference have become popular in performing genetic variation studies. Massively parallel next generation sequencing technology was used to resequence the chicken genome in an effort to identify loci associated with chicken domestication (Rubin etal. 2010). Rubin et al. generated 44.5-fold coverage of the chicken genome using pooled genomic DNA representing 8 different chicken populations, as well as the ancestral line, the Red Jungle Fowl. Results from this study identified over 7,000,000 SNPs, almost 1,300 deletions and some selective sweeps. A selective sweep occurred at the locus for thyroid stimulating hormone receptor (TSHR), playing a crucial role in metabolic regulation and photoperiod, was identified as a genetic marker indicative of domestication.

Genome wide association studies (GWAS) have been used to find differences in certain regions of the chicken genome under selection. Lan et al. (2015) used the GWAS method to isolate specific loci between the broiler and layer chickens to support the production of indigenous chicken breeds. Results suggested that alternative indigenous breeds could be used in place of the broiler and layer chicken. Next generation sequencing and GWAS methods have also been utilized to investigate genetic resistance and susceptibility to certain diseases. Two domestic chicken lines, divergently selected for susceptibility to Marek's disease (MD), were used to identify the genetic and molecular basis for resistance to MD. Yan et al., reported that birds found to be susceptible to MD had fixed copy number variation regions (CNVRs) containing several genes involved in the MAPK signaling pathway and 62 genes strongly associated for the susceptibility to MD. The divergently selected Arkansas Progressor (AP) and Arkansas Regressor (AR) chicken lines can be an animal model for disease resistance and investigation on genome wide scale variation, for AP and AR lines may provide genetic information that can be used for diagnoses,

treatment, and genome-wide selection strategy. GWAS can identify genome variation and provide insight to the genetic correlates of phenotypic changes.

The objective of this study was to identify potential genetic biomarkers for tumor regression and tumor progressions using the two divergently selected AP and AR chicken lines.This was the first whole genome sequencing conducted to analyze SNP variations in these two chicken lines. Pooled genomic DNA representing both lines was sequenced using Illumina technology and a reference based assembly using the sequences of Red Jungle Fowl, common ancestor of all domestic chicken breeds. Not only can this study provide insight into genetic variation between the AP and AR lines, but also provide useful data toward domestication and biomedical research.

MATERIALS AND METHODS

Genetic Lines

Two divergently selected chicken lines, AP and AR, which have been maintained by N.B. Anthony at the Arkansas Agricultural Experimental Station (AES), were use in this study. AP line chickens are susceptible to tumor progression after infection of RSV, while AR line chickens are resistant to tumor development after RSV infection. To produce the AR and AP birds for this study, matings were achieved via artificial insemination of 148 females with pooled semen from 74 males. Details on the method of selection used to create these two lines are described in Chapter 1.

Sampling and DNA preparation

To obtain DNA samples for whole-genome sequencing, 0.5 ml of blood was collected via wing web puncture from 12 birds each from the AP and the AR lines. Genomic DNA was then extracted using a QiaAmp DNA mini kit (Qiagen, Valencia, CA) following manufacturer's instructions. Extracted DNA quantity was determined by NanoDrop spectrophotometry (Thermo Scientific, Waltham, MA) and quality by agarose gel electrophoresis. Equal nanograms of DNA from 10 samples showing the highest quality were pooled to represent each respective line. Quality of DNA was based on DNA concentration.

Genome Sequencing and Assembly

Library construction and whole-genome sequencing of the pooled DNA were performed by the National Center for Genome Resources (NCGR: Santa Fe, NM) using Illumina HiSeq 2x100 bp paired end read technology. The raw sequence read data were assembled and aligned with the Red Jungle Fowl (gal gal 4; GenBank accession number: GCA_000002315.2), using SeqMan NGen program of the Lasergene software package (version 12.0 DNAstar, Madison, WI). Assembly parameters are listed in Table 1.

Data Analysis I (SNP filtering)

After assembly, SeqMan Pro (LaserGene software program) was used to identify candidate SNPs for the entire genome. The SeqManPro SNP report was exported to Microsoft Excel and further analyzed. Duplicate SNPs homozygous in both AP and AR, were filtered out leaving only unique SNPs to observe any fixations between the two lines. The SNP percentages, or SNP frequencies, are defined as the ratio of reads containing the SNP to total read number. Any SNPs with less than 75% SNP frequency and a read depth of less than 10 were filtered out of the dataset to obtain more reliable SNPs. The SNP frequency of the AR line was subtracted from those of the AP line. If frequency in AP- frequency in AR was equal to – 1 (1.00), the SNP was to be fixed at 100% frequency for the AR line at that SNP location. Likewise, when the difference of SNP frequency between AP and AR is ±1.00, the SNP was considered to be fixed at 100% in the AP line. These were considered potential areas of the genome to further analyze to find candidate genes containing biomarkers.

Using the remaining SNP data for all chromosomes, SNP frequencies were averaged over a 40-SNP window to show determine if any clusters of SNPs were present at chromosomal locations. This data was then graphed to identify clusters of fixed SNPs at their chromosomal location (data not shown). The outlying peaks which are much larger than the normal distribution of the graph are the areas of clustered SNPs. These areas were targeted for initial SNP analysis and further analyzed to find genes with functions associated with tumor growth, cancer, and the cell cycle.

Data Analysis II (SNP filtering)

For initial genomic analysis of SNP data, SNPs showing a depth of 10 or higher were considered for validation (Data Analysis I). The validation results revealed that these SNPs were false positives caused by filtering out important SNPs. Thus, the depth limit was decreased to three rather than ten. Duplicated SNPs with a read depth of two or less were filtered out to ensure fixation of that SNP in only one of the two lines. The unique SNPs with a SNP frequency less than 75% were then filtered out of the dataset High frequency, unique SNPs found in coding DNA sequence (CDS) and non-coding regions were grouped into various categories including: mutations of frameshift, no start, no stop, non-synonymous, and non-sense, which may generate mutations in the amino acid sequences which may have an impact on the phenotypic changes. SNPs that appeared to be fixed were confirmed further at the SNP locations using the alignment view of SeqMan Pro program. The actual reads were aligned on the references where the SNP frequency in each line could be verified. SNPs lying within genes with high frequencies and also functional activities in the cell cycle, cancer, the immune system or tumor growth were chosen to further investigate in a larger population.

RESULTS AND DISCUSSION

Genome Sequencing and Assembly

The average coverage for all chromosomes was 9.9x. The total number of reads for AP and AR lines was 69,221,284 and 55,368,344 respectively (Table 2). SeqMan Pro identified more than 9,000,000 SNPs in both AP (4,948,899) and AR (4,407,630) lines, most SNPs being in chromosomes 1-3. Figure 2 shows the distribution of SNPs per chromosome.

Data Analysis I (SNP filtering)

During the data filtering process in Microsoft Excel, over 388,000 SNPs were found with a read depth of 10 or more in the AP and AR lines (Table 2), but only 12,000 were unique between the AR and AP lines. Further filtering revealed that 7,162 unique SNPs had a 75% or higher SNP frequency.

Although SNPs located within non-protein coding regions can be significant biomarkers, not all SNPs could be considered in this study due to the time required for a large scale SNP analysis. If a candidate SNP outside the protein coding region is found to be highly associated with the phenotype of interest, it is plausible that the association is due to linkage disequilibrium. However, the linkage between two markers would decrease over time due to meiotic recombination events eventually breaking the linkage of the marker and phenotype (Brookes, 1999). Although this is true, it is still possible for SNPs in non-coding regions of genes to be associated with the phenotype of interest.

Once filtering was completed, ten SNPs were chosen based on SNP frequency and the gene function in which the SNP resided. The ten candidate genes chosen to test included RALA, STK17A, CDKN2A, CDKN2B, POU6F2, NCK2, PSMA2, SMOC2, ADRBK2, and CDK13.The gene descriptions are listed in Table 4. As mentioned before, these candidate SNPs were shown to be false positives after testing these SNPs in a larger number in AP and AR chickens. Thus the second data analysis was performed as follows.

Data Analysis II (SNP filtering)

After the second data analysis, candidate SNPs were selected, and the nine new SNPs were found to be reliable candidate SNPs for further testing. All had a SNP frequency of 100% in the initial genomic analysis from the 10 pooled samples. After the second filtering process was completed, the unique SNPs were categorized by mutation: Frameshift (134), No Start (25), No Stop (7), Non-synonymous (6884), and Nonsense (112) (Table 6). The second data analysis revealed 9 candidate genes containing a SNP of interest. These genes include: IL16, IL2RA, CDK13, CASP9, FGF14, MAP1A, POU6F2, NCK2, and ATAD5. The descriptions of these genes can be found in Table 5.

Genome Sequencing and SNP Analysis

Whole genome sequencing has been proven to be a reliable method in identifying loci under intense genetic selection in chickens as well as other species. This method has been performed to study the domestication of the chicken, the impact of genetics on disease resistance, and marker assisted selection in agricultural chicken populations (Rubin et al., 2010; Lan et al., 2015). Gheyas et al. (2015)

detected 15,000,000 SNPs using next generation sequencing in diverse chicken populations in order to predict variants with potential functional implications from coding and non-coding regions. The study reported 183K SNPs, which resulted in amino acid alterations and identified regions of local fixation within commercial broiler and layer lines. Ninety-four variants were detected from various putative selective sweeps to be fixed exclusively in broilers, most of which resided on chromosome 1 and represented genes: PARPBP, PMCH, NUP37, CCDC53, GNPTAB, SYCP3 and CHPT1. The nine SNPs detected in AP and AR lines in this study have the potential to indicate RSV- induced tumor susceptibility and may provide insight into the etiology of cancer causing viruses such as the Human Papilloma Virus (HPV), infecting about 14 million people each year resulting in 11,000 women getting cervical cancer (Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2015). The whole genome sequencing results from this study can also provide potential opportunities to further study the genetic mechanisms involved in tumor development in humans. The effects of intense human-driven selection based on phenotype can also be studied to aid in animal breeding programs.

Today, next-generation sequencing plays a major role in clinical applications of genomic disorders, especially cancer. Various types of genetic mutations, point mutations, genomic rearrangements, gene fusions, or chromosomal alterations can contribute to disease initiation and/or progression (Kim et al., 2013). The AP and AR chicken lines serve as an excellent model to study the effects of genetic mutations which may contribute to tumor growth. The genetic mutations found to be responsible for disease can be used as biomarkers for prognosis or treatment therapies in humans. Roychowdhurty et al. (2011) developed a cancer patient protocol that includes genome sequencing of the cancer genome to evaluate the mutation profiles of the patient to aid in clinical decisions. In their pilot study, sequencing tumor board (STB) found that a patient with metastatic colorectal cancer had an NRAS mutation and CDK8 amplification. This finding concluded that BRAF/MEK inhibitors and PI3K and/or CDK inhibitors could be beneficial for that patient. Some of the same markers found in the human genome have a similar counterpart in the chicken genome, giving reason to use the chicken as a model for biomedical research.

CONCLUSION

The chicken genome has 20,000- 23,000 genes and 1 billion base pairs and the human genome has 20,000- 25,000 genes and 3.2 billion base pairs. The chicken and human genomes contain 60% corresponding genes, which are 75% identical on average, making the chicken a useful model for genome evolution and human biology (National Institutes of Health, 2004). Genome analysis showed that the puffer fish shares 72% of the corresponding chicken and human genes, suggesting similar percentage of gene correspondence in all vertebrates. The SNP data for AP and AR chicken resulting from this thesis can be further analyzed for cellular interaction pathways, contributing knowledge of the genetic mechanisms involved immune response to disease. All nine of the selected SNPs were located within a gene with functionality related to various signaling pathways in cells and have a corresponding gene in the human genome. The chicken is a great model for human cancer research, as well as other diseases. Other chicken lines at the University of Arkansas have been created not only to improve economic value but also to serve as models for various diseases, such as hypertension and vitiligo in humans (Anthony, N.B.; Erf, G.).

REFERENCES

Amaral, Andreia J., Luca Ferretti, Hendrik-Jan Megens, Richard P. M. A. Crooijmans, Haisheng Nie, Sebastian E. Ramos-Onsins, Miguel Perez-Enciso, Lawrence B. Schook, and Martien A. M. Groenen. (2011). Genome-Wide Footprints of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA. *PLoS ONE*

Brookes, Anthony J. (1999). The Essence Of SNPs. *Gene*. 177-86.

Genital HPV Infection - Fact Sheet Centers for Disease Control and Prevention. 23 Feb. 2015. Centers for Disease Control and Prevention. Web. 16 Nov. 2015.

Gheyas, A. A., C. Boschiero, L. Eory, H. Ralph, R. Kuo, J. A. Woolliams, and D. W. Burt. (2015). Functional Classification of 15 Million SNPs Detected from Diverse Chicken Populations. *DNA Research*. 205-17.

Kim, T., Lee, S. and Chung, Y. (2013). Clinical applications of next-generation sequencing in colorectal cancers. *World Journal of Gastoenterology*. 19(40): 6784-6793.

Lan, D., Hu, Y. D., Zhu, Q., Li, D. Y., and Liu, Y. P. (2015). Verification of specific selection SNPs between broiler and layer chicken in Chines indigenous chicken breeds*. Genetics and Molecular Research.* 14(3): 8388-8396.

Researchers Compare Chicken, Human Genomes. 2004. Researchers Compare Chicken, Human Genomes. National Institutes of Health and National Human Genomic Research Institute, Web. 16 Nov. 2015.

Roychowdhury, S., Iyer, M. K., Robinson, D. R., Lonigro, R. J., Wu, Y., Cao, X., Kalyana-Sundaram, S., Sam, L., Balbin, O.A., Quist, M. J., Barrette, T., Everett, J., Siddiqui, J., Kunju, L. Pl, Navone, N., Araujo, J. C., Troncoso, P., Logothetis, Talpaz, M. and Chinnaiyan, A. M. (2011). Personalized oncology through integrative high-throughput sequencing: A pilot study. *Science Translational Medicine.* 3(111): 111ra121.

Rubin, C., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman,T. Sharpe, S. Ka, F. Hallböök, F. Besnier, O. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen,P. Siegel, K. Lindblad-Toh, and L. Andersson. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature.* 464:587-591.

Yan, Y., Yang, N., Cheng, H. H., Song, J., and Qu, L. (2015). Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. *BMC Genomics.* 16(1): 843.

LEGEND OF TABLES AND FIGURES

Table 1: Assembly parameters

Assembly parameters for the alignment of AP and AR genomes to the reference genome. SeqMan NGen program of the Lasergene software package (DNAstar, Madison, WI) was used for the assembly with listed parameters

Table 2: Combined summary for NGS of AP and AR lines

Combined total numbers of contig length, contig length without gaps, total sequence length, number of sequences, average coverage, and cumulative contig length (without gaps) for each chromosome for AP and AR genomes.

Table 3: Summary for NGS results for AP and AR lines

Total number of reads, aligned reads, and unaligned reads for AP and AR chicken lines

Table 4: Data Analysis I- gene descriptions (RefSeq, Jul 2008)

Gene names, descriptions, and SNP locations for initial ten SNPs chosen as candidate biomarkers resulting from the first data analysis. SNPs listed were validated as false positives except NCK2 and POU6F2

Table 5: Data Analysis II- gene descriptions (RefSeq, Jul 2008)

Gene names, descriptions, and SNP locations for nine SNPs chosen as final candidate biomarkers resulting from the second data analysis

Table 6: Summary of categorized mutations after Data Analysis II

Number of categorized mutations after second data analysis for AP and AR lines

Table 7: Candidate SNPs chosen for potential genetic biomarkers

Gene names, reference position, reference base, mutation type, called base, SNP frequency, genotype, and feature of nine final candidate SNPs chosen for potential genetic biomarkers

Figure 1: Possible number of total SNPs per chromosome for AP and AR lines

Possible number of total SNPs per chromosome for AP and AR lines combined, with more than 4 million SNPs occurring in chromosomes 1-3.

Table 1: Assembly parameters

Layout Options	
Repeat read placement	place once
Maximum repeat count:	100
Maximum total read:	10000000
Assembly output format	BAM assembly package
Alignment Options	
Minimum aligned length	25
Maximum gap size	6
Minimum match percentage	93
Mismatch penalty	20
Gap penalty	30
Auto trim reads	on
SNP Options	
Calculate SNPs	on
SNP calculation method	Diploid Bayesian
Minimum SNP percentage	5
SNP confidence threshold	10
Minimum SNP count	2
Minimum base quality score	5
Check strands	off

Table 2: Combined summary of NGS for AP and AR lines

Chromosome	Contig Length	Contig Length(woGaps)	Total Seq. Length	Number Seq.	Avg. Coverage	Cumulative Contig Length(woGaps)
mtDNA	16795	16775	250100	2501	14.89	16775
1	195486522	195276750	2009697400	20096974	10.28	195293525
2	148964650	148809762	1509047700	15090477	10.13	344103287
3	110564444	110447801	1131729100	11317291	10.24	454551088
4	90317033	90216835	925996300	9259963	10.25	544767923
5	59644960	59580361	613492100	6134921	10.29	604348284
6	34991399	34951654	364903100	3649031	10.43	639299938
7	36284861	36245040	379173800	3791738	10.45	675544978
8	28798905	28767244	299465000	2994650	10.4	704312222
9	23468928	23441680	246802800	2468028	10.52	727753902
10	19933620	19911089	206727600	2067276	10.37	747664991
11	19423623	19401079	203894200	2038942	10.5	767066070
12	19919731	19897011	208543100	2085431	10.47	786963081
13	17780811	17760035	184949900	1849499	10.4	804723116
14	15179679	15161805	159867500	1598675	10.53	819884921
15	12670337	12656803	131905300	1319053	10.41	832541724
16	535984	535270	4881900	48819	9.11	833076994
17	10468340	10454150	123699000	1236990	11.82	843531144
18	11232318	11219875	111650900	1116509	9.94	854751019
19	9994848	9983394	105255300	1052553	10.53	864734413
20	14318198	14302601	146270200	1462702	10.22	879037014
21	6811114	6802778	69697900	696979	10.23	885839792
22	4085298	4081097	43324600	433246	10.61	889920889
23	5729945	5723239	57035500	570355	9.95	895644128
24	6330974	6323281	64356500	643565	10.17	901967409
25	2193591	2191139	16783800	167838	7.65	904158548
26	5336208	5329985	49848200	498482	9.34	909488533
27	5215342	5209285	49829200	498292	9.55	914697818
28	4747584	4742627	41466400	414664	8.73	919440445
32	1070	1028	451000	4510	421.5	919441473
W	1249589	1248174	10306300	103063	8.25	920689647
Z	82420785	82363669	473442700	4734427	5.74	1003053316
O	966208	965146	8149000	81490	8.43	1004018462
P	801117	799899	10439700	104397	13.03	1004818361
totals	1005884811	1004818361	9963333100	99633331		
average	29584847.38	29553481.21	293039208.8	2930392.088	22.21647059	

Table 3: Summary of NGS Results for AP and AR lines

LINE	Total # of reads	Total # of reads aligned	Total # of reads not aligned
AP	69,221,284	55,224,050	10,903,306
AR	55,368,344	44,328,649	8,551,173

Table 4: Data Analysis I- gene descriptions (RefSeq, Jul 2008)

Gene	SNP location (Chr:Mpb)	Gene descriptions
RALA	2:504442109	v-ral simian leukemia viral oncogene homolog A (ras related)
STK17A	2:5211119	Serine/threonine kinase 17a
CDKN2A	Z:78453153	Cyclin-dependent kinase inhibitor 2A
CDKN2B	Z:78463164	Cyclin-dependent kinase inhibitor 2B
POU6F2	2:50117844	POU class 6 homeobox 2
NCK2	1:134690198	NCK adapter protein 2
PSMA2	2:51749203	Proteasome (prosome, macropain) subunit, alpha type 2
SMOC2	3:40985084	SPARC related modular calcium binding 2
ADRBK2	15:7043072	Adrenergic, beta, receptor kinase 2
CDK13	2:50507153	Cyclin-dependent kinase 13

Table 5: Data Analysis II- gene descriptions (RefSeq, Jul 2008)

Gene	SNP location	Gene descriptions
CDK13	2:50469650	Cyclin-dependent kinase 13
NCK2	1:134690198	NCK adapter protein 2
POU6F2	2:50117844	POU class 6 homeobox 2
IL2RA	1:3380404	Interleukin 2 receptor, alpha
ATAD5	18:6611691	ATPase family, AAA domain containing 5
CASP9	21:4919322	Caspase 9, apoptosis-related cysteine peptidase
FGF14	1:142233297	Fibroblast growth factor 14
IL16	10:11603087	Interleukin 16
MAP1A	10:19701184	Microtubule-associated protein 1A

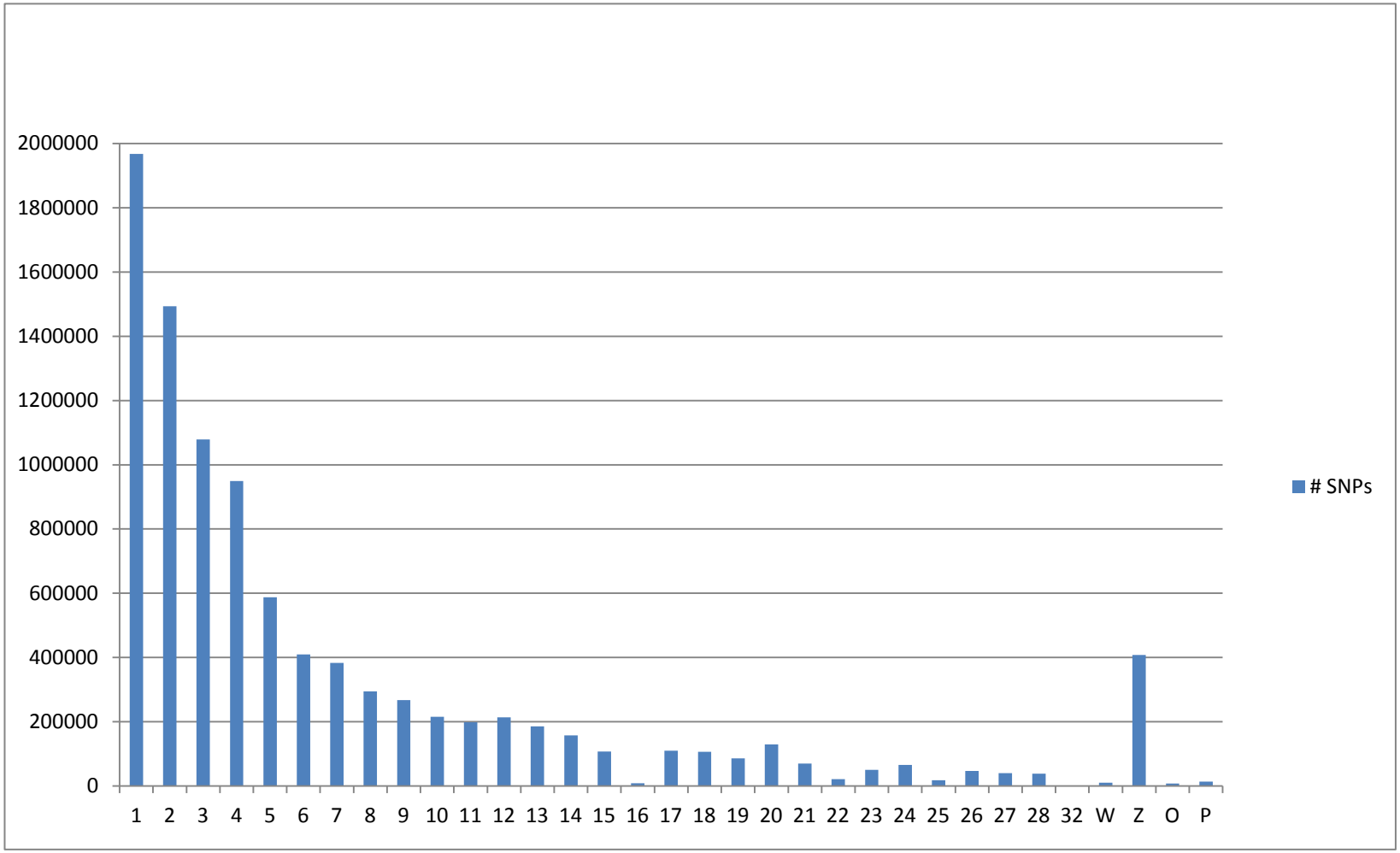
Table 6: Summary of categorized mutations after Data Analysis II

Type of SNP	# SNPs in AR	# SNPs in AP
Non-synonymous	5586	6771
Non-sense	45	66
No Start	19	20
No Stop	7	7
Frameshift	88	148

Table 7: Candidate SNPs chosen for potential genetic biomarkers

GENE	LINE	CHR	REF POSITION	REF BASE	MUTATION	CALLED BASE	SNP % of called base	Genotype	Feature
IL16	AP	10	11603087	T	Non-synonymous	C	100%	Homo. Var	CDS
FGF14	AP	1	142233298	T	Non-coding region	C	100%	Homo. Var	gene
IL2RA	AP	1	3380404	C	Nonsense	T	100%	Homo. Var	CDS
MAP1A	AP	10	19701184	G	Non-synonymous	A	100%	Homo. Var	CDS
CDK13	AR	2	50496950	G	Non-synonymous	A	100%	Homo. Var	CDS
ATAD5	AR	18	6611691	C	stop codon	T	100%	Homo. Var	CDS
CASP9	AR	21	4919322	G	Nonsense	A	100%	Homo. Var	CDS
POU6F2	AR	2	50117844	T	Non-coding region	G	100%	Homo. Var	gene
NCK2	AR	1	134690198	T	Non-coding region	T	100%	Homo. Var	gene

Figure 1: Possible number of total SNP's per chromosome for AP and AR lines



CHAPTER 3: VALIDATION OF CANDIDATE SNPs ASSOCIATED WITH RSV-INDUCED TUMOR DEVELOPMENT

INTRODUCTION

Identifying and understanding genetic mechanisms by which evolutionary changes occur can provide knowledge to better understand the association between genotypic and phenotypic characteristics of organisms. Domestic animals, such as the chicken, serve as excellent models for genetic studies involving phenotypic evolution caused from disease. Human-driven selection has led to genetic adaptations responsible for significant phenotypic changes. Current advancements in molecular genetics have driven the identification of genetic markers and provided opportunities to improve conventional selection when the desired phenotype is not measurable and/or has low heritability (Dekkers, 2003).

The poultry industry has struggled to link targeted phenotypes with genotypes, but continues to research as more advanced molecular methods are discovered. GWAS (Genome Wide Association Studies), differential gene expression, and resequencing are among the commonly used methods for genetic variation studies. Luo et al. (2014) performed a GWAS using 43,211 SNP markers to identify the loci involved in the immune response to the infection of infectious bronchitis (IBV) in the chicken. The study detected 20 significant SNPs related to the anti-body response to IBV. Differential expression between Marek's disease (MD) infected spleens and non-infected spleens results showed down-regulation of gga-miR-199-3p at 14 and 28 dpi and expression of miR-140-3p decreased at 14 dpi, suggesting an important role in the early stages of tumorigenesis (Lian et al., 2015). Huang et al. (2015) used DNA sequencing and PCR-RFLP as a means to link genotypes of BMP15 and GDF9 to egg production. Two SNPs, one in each gene, were found to be associated with total egg production at 300 d. of age and age of first lay. This study concluded that Marker Associated Selection (MAS) of BMP15 genotypes could significantly improve total egg production at 300 d. of age in Shaobo hens. Whole genome sequencing can also be utilized to identify genetic mutations. This method has been used to pinpoint functional variants from association studies, improve evolutionary biology research, hence creating the ability to predict disease susceptibility and drug response.

Whole genome sequencing can help researchers understand the functionality of the genome in relation to the development of genetic adaptations. Before the development of whole genome sequencing, Restricted Fragment Length Polymorphism (RFLP) analysis and microsatellite analysis were

common techniques used to find genetic variation. These methods investigate small regions of the genome with minimal knowledge of the gene sequences. Whole genome sequencing has become less time-consuming and costly, providing benefits to genetic research in terms of efficiency and quality. In this study, the Arkansas Progressor (AP) and Arkansas Regressor (AR) chicken lines were used as a model to identify genetic mutations responsible for Rous sarcoma virus (RSV) induced tumor development. Whole genome sequencing was performed and genome-wide SNPs were identified and analyzed for AP and AR lines. Nine SNPs were identified as being potential biomarkers for tumor regression or progression in chickens post RSV inoculation. The nine candidate SNPs were further tested in larger populations to confirm their presence in AP and AR lines. SNPs showing high frequencies and divergence in larger populations were chosen as genetic biomarkers for marker-assisted selection for RSV tumor-induced resistance.

MATERIALS AND METHODS

DNA sampling and preparation

At two weeks of age, blood was sampled via wing web puncture from more than 50 birds for each the AP and AR chicken lines. Approximately 2-3 drops (100-200µL) of blood was collected and stored in 100µl sodium citrate ($C_6H_8O_7$) to prevent clotting. Genomic DNA was extracted from the blood using the Wizard SV 96 Genomic DNA Purification System (Promega; Madison, WI) following manufacturer's instructions with some modifications. In the PCR plate format, the top four rows (48 wells) contained AR line samples and the bottom four rows (48 wells) contained AP line samples. Whole blood (10µL) was carefully mixed with proteinase K (6.7µL/µL) and incubated at room temperature for 10 minutes. The lysis buffer (300µL) and triton X-300 (33µL) was mixed and then added to RNase A (55µL) in one boat. This mixture was then added to the blood/proteinase K mixture and blood clots were dissociated by repetitive pipetting. After complete dissociation of blood clots, a vacuum manifold system was used to filter lysate through the filters of the binding plate. Three consecutive cycles of ethanol washing was applied and followed by elution with nuclease- free water (200µL). DNA concentration was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). For all samples, a 1ng/µL dilution was prepared in empty 96 well plates as an aliquot for PCR.

Primer Design

A forward, reverse, and sequencing primer were designed with reference to the RJF genome sequences (gal gal 4;GenBank accession number: GCA_000002315.2) at the SNP locations using online program Primer 3 Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) (Untergrasser et al., 2007). Primers were designed to have 20-22 nucleotide bases and approximately 50% GC content. Sequencing primers were designed to anneal to at least 50bp or more upstream of the SNP location. Forward and reverse primers were designed to anneal to the flanking regions of the sequencing primer and the SNP location. The lengths of PCR product ranged between 431-520bp. All working primers were synthesized by Integrated DNA Technologies (Ames, IA). Working primers were prepared to a concentration of 20ng/µL. A total of 9 primer sets were designed, one for each candidate gene, and these can be viewed in Table 1.

PCR

PCR was performed on a 96-well format for all 9 candidate genes for both the AR and AP lines. These included IL16, FGF14, IL2RA, MAP1A, CDK13, ATAD5, CASP9, POU6F2, and NCK2. The top 4 rows of the 96-well plate contained AR samples and the bottom 4 rows contained AP samples. The 96 well format can be seen in Figure 1. Total reaction volume for PCR was 25µL: 5µL DNA (1ng/µL), 2.5µL 10x buffer, 1µL 2.5mM dNTP mix, 2µL F primer, 2µL R primer, and 0.5µL Taq Polymerase (NEB; Ipswich, MA). PCR conditions were as follows: denaturation at 95ºC for 1 min, 40 cycles of amplification (95ºC for 30 sec, 60ºC for 1 min, 72ºC for 1 min), and a final extension at 72ºC for 10 mins. PCR assays were performed using Applied Biosystems 2720 Thermal Cycler (Life Technologies, Carlsbad, CA). Validation of PCR product was performed by 1% agarose gel electrophoresis.

Gel Electrophoresis

To test quality of PCR product, 5µL was mixed with 6x loading dye, prepared with bromophenol blue and xylene cyanol dyes. The mixtures were loaded into wells of a 1% agarose gel using 1x TAE buffer solution. The first well was loaded with 5µL Hi-Lo DNA molecular weight marker (Bionexus, Inc.,

Oakland, CA) to confirm PCR product size of each sample. Gel images were obtained using a Bio-Rad Gel Doc XR imaging system (Bio-Rad Laboratories, Hercules, CA).

PCR product purification

After gel electrophoresis, PCR product was purified using the Wizard SV 96 PCR Clean-Up System (Promega; Madison, WI) following manufacturer's protocol. Prior to PCR purification, 8µl of four different PCR products were pooled into one 96-well plate totaling 32µl in each well. Each well contained four different samples derived from the same bird (Figure 2). Before pooling these samples, the cross-specificity of the sequencing primers to the other PCR product sequences was assessed using the BLAST function (NCBI). Membrane-binding solution (32µL) was added to the pooled PCR products and vacuum pressure was applied. Three washing steps with 95% ethanol were applied and the PCR products were eluted with 100µL nuclease-free water. The DNA concentration was quantified using the NanoDrop 1000.

Sanger Sequencing

After DNA concentration was quantified, further pooling was performed. From the 96 well plates, 3µL from each well of each row were pooled into one tube (i.e.rows A1-A12). This resulted in 8 different pooled samples containing purified PCR products of 4 different genes. Because these pooled samples contained PCR products for 4 different genes, four separate aliquots were made for each sequencing primer. Each sequencing primer was added separately so as to only show results for the desired gene. To prepare for sequencing, 20ng of DNA was mixed with 1.5µl of the sequencing primer and distilled water to yield 13µl total for sequencing reaction submission. Sequencing was performed in the DNA Core Lab at the Tyson Center of Excellence for Poultry Science, University of Arkansas (Fayetteville, AR). Sequencing results were analyzed with Sequence scanner software v. 1 (Applied Biosystems of Thermo Fisher Scientific, Waltham, MA). Peak heights at SNP locations were examined and the ratios of allele frequencies at the SNP locations were recorded.

RSV Inoculation and blood collection

The bird population being tested was also observed for phenotypic data collection. Approximately 100 birds from each line were hatched and maintained at the AES (Fayetteville, Arkansas). To produce the AR and AP birds for this study, matings were achieved via artificial insemination of 148 females with pooled semen from 74 males. All chicks were given a wing band with a unique identifying number and then placed randomly in pens. Feed and water were provided *ad libitum* during the grow-out of the birds. Phenotypic data (tumor scores) were recorded for each bird/wing band in order to see tumor growth for individuals over a duration of time.

At six weeks of age, all birds were inoculated subcutaneously with 70pfu Bryan-titer strain Rous Sarcoma Virus (0.1mL) in the opposite wing web from where the blood sample was taken. The RSV virus sample used in this procedure was of the C RSV subgroup (specifically RAV-49), and came from Dr. Aly Fadly, USDA-ARS, Avian Disease and Oncology Laboratory, East Lansing, MI. The birds' tumors were monitored three times a week and recorded for 8-10 weeks post inoculation to track regression or progression of tumors. A sizing scale from 0-5 was created to record the different stages of tumor growth. A bird was considered resistant if no tumor developed on the wing web after inoculation. If a tumor of any size developed, and regressed to zero, it was considered completely regressed. If a tumor of any size developed, and regressed to 1, it was considered more resistant to RSV than birds of higher scores. Any bird that developed a tumor without any regression was considered completely progressed. If a bird developed a tumor and it regressed to anything greater than 1, it was considered less resistant to RSV. Sizing scale can be reviewed in Table 2.

Images of example tumor sizes are shown in Figures 3-9. Once a tumor reached a score of 5 for one week, and showed no sign of regression, the bird was humanely euthanized. Also, if the skin covering the tumor broke open, and the contents were visible outside the tumor, the bird was humanely euthanized. All birds were humanely euthanized 10 weeks post inoculation (AVMA Guidelines).

Phenotypic Data Collection

The tumor scores for each line were recorded every 3-4 days on average for 67 days. The same 48 birds of each line used for SNP testing were used for phenotypic analysis. The tumor scores were averaged for each day and then graphed against the days post inoculation (dpi). These data can be seen in Figures 14 and 15.

RESULTS AND DISCUSSION

Genotypic Results

To determine SNP frequencies in large samples sizes for AP and AR lines, pooled PCR products from 12 chickens were subjected to sequencing and the peak heights examined at SNP locations was used to measure genotypic ratios. This method can illustrate an estimation of SNP frequencies in large pooled samples. The accuracy of allele frequencies of pooled PCR products was determined by the comparison of 12 individually sequenced samples at the SNP locations. Comparing peak heights of pooled samples to determine the SNP frequency has been confirmed recently (Lee et al., 2014).

Results from PCR and sequencing of the first set of candidate genes derived by data analysis I revealed that candidate SNPs were false positives, and only SNPs present in genes NCK2 and POUF6 occurring in the AR line showed fixed segregations from the AP line (data not shown). Although candidate SNPs showed higher frequencies in the counter line, the fixation may not be high enough to be considered a candidate gene, since this SNP was seen in both lines. Therefore, data analysis II was performed and the seven new candidate SNPs, in addition to the SNPs occurring in NCK2 and POUF6, were then tested in the same population of birds, 48 chickens per line.

The second set of candidate SNPs was tested and results were more promising; the sequencing results are listed in Table 3. The SNP, located in the IL-16 (interleukin-16) gene, resulted in 100% complete divergence between the two lines. The AR line had the same base (T) as the reference genome with a 100% frequency and the AP line had a different base (C) than the reference genome with a 100% frequency. In chickens, IL-16 is an inflammatory cytokine and also a precursor of protein for pro-IL-16 (Min et al., 2004). Yellapa et al. (2012) reported that higher levels of IL-16 serum were associated with

tumor growth resulting from tumor-associated neoangiogenesis (TAN) in laying hens. The serum level frequency of IL-16 was significantly higher in tumor-bearing hens than hens without tumors. Results concluded that serum levels of IL-16 may potentially be indicative of TAN. IL-16 is a cytokine that is involved in the innate immune response. A cytokine can play a major role in the regulation of both innate and adaptive immunity. More specifically, most interleukins are responsible for promoting the development and differentiation of T, B lymphocytes or natural killer cells (NK). Because the AP line is known for tumor progression, The SNP found in IL-16 could play a role in tumor development and resistance to RSV.

A SNP in the CDK13 (cyclin-dependent kinase 13) gene had complete divergence between the two lines as well, however only 36 out of the 48 samples per line had successful sequencing results. Sequencing results showed that this SNP was 100% fixed in the AR line with the “A” base, while the AP line was fixed at 93.3% with the “G” base. Ansari et al. (2015) discovered genes which produced proteins linked to pancreatic cancer were unique to pancreatic cancer patients and CDK13 was included in the list. CDK13 , also known as CDCL25 (cell division cycle 2- like 5), has also been found to have high oncogenic activity linked to tumorigenesis and was also amplified by an average of 30% in human patients with hepatocellular carcinoma or colorectal cancer (Kim et al., 2012). CDK13 has been identified as an essential protein that influences the cell cycle by regulating the activation of cyclins. Tumor development is thought to result from a genetic alteration of a protein, which then affects the proliferation of a single cell. That cell continues to proliferate until there is a tumor clonal population. Additional mutations such as the SNP found in CDK13 may affect the cell cycle and cause overactive cell proliferation leading to tumor development such as the sarcomas induced by RSV. Overactive cell proliferation is considered an abnormality in the cell cycle and can be controlled by genetic mutations (Cooper, 2000).

IL2RA (interleukin-2 receptor, alpha) had a 100% SNP (T base) frequency in the AP line and an 85.5% SNP (C base) frequency for the AR line. This gene, aka CD25, is the alpha chain of the interleukin receptor, which is found on activated T cells. In humans, this gene was shown to have lower expression

in HIV infected individuals, possibly contributing to immune deregulation (Mendez-Lugares et al., 2014). A mutation in the IL2RA gene could help identify the mechanism of virus-induced diseases such as HIV.

CASP9 (caspase-9) also had promising results having a 100% SNP (G) frequency for the AP line, with the same base as the reference. The AR line did not match the reference and had an 83% SNP for base A and a 17% SNP frequency for base T. CASP9 was discovered to play a role in the direct regulation of apoptosis, and in turn tumor development. CASP9 is an oncogenic gene which is activated by p53, a tumor suppressor gene, to promote apoptosis (Sonegas et al., 1999). CASP9 is yet another candidate gene that could be contributing to the tumor development in the AR and AP lines due to its role in tumor development in humans.

Additionally, ATAD5 (ATPase family, AAA domain containing 5), showed a 100% SNP (C base) frequency in the AP line matching the reference base while AR had an 83.3% SNP (T base) and 17% SNP (A base) frequencies. Bell et al. (2011) discovered that this gene may be important in suppressing genomic instability in mammals *in vivo*. Using mice as their model, they found that 90% of ATAD5 haploinsufficient mice developed tumors. This may indicate that a mutation in ATAD5 affects tumor development. The SNP found in ATAD5 for the AR line is different from the reference, suggesting that the AR line may have experienced a mutation in favor of tumor regression.

SNP frequency in FGF14 (fibroblast growth factor 14) in the AP line was 100% (C base) while the AR line showed 62.5% (T base) SNP frequency and matched the reference base. Although this SNP is not completely divergent between the two lines, the 100% in the AP line may be significant. This SNP is also located in the intronic region; however, it is possible for a mutation in an intronic region to affect protein synthesis, especially if the mutation is close to the splicing site (Harland et al., 2001). The FGF protein family (fibroblast growth factors) are responsible for binding to the FGF receptors to create a dimerization to activate tyrosine phosphorylation in the SHP2/RAS/ERK signaling pathway ultimately creating proteins (Goldfarb, 2001). Tyrosine phosphorylation is an essential reaction in cell differentiation, the cell cycle, signal transduction and gene regulation and transcription (Abel et al., 2012; Meijer et al., 1991; Lin et al., 2006). If a mutation caused an FGF protein to be non-functional, tyrosine phosphorylation

may not be activated and potentially affecting protein synthesis, which in turn could affect immune response.

Another candidate gene showing favorable results was MAP1A (microtubule-associated protein 1A) with a 97.5% SNP (G base) frequency in the AR line (matching reference base) and a 60% SNP (A base) frequency in the AP line. This gene was found to have high cytoplasmic expression in humans with gliomas (Meyer, 2014) and MAP1A mRNA affected tumor size in rats when delivered straight into the tumor via a mini-osmotic pump (Matsuno et al., 2004). Meyer's (2014) results indicate that MAP1A may have a connection to tumor development and growth, making it a potential gene involved in the development of RSV-induced tumors in the AR and AP chicken lines.

POU6F2 (POU class 6 homeobox 2) SNP frequencies were 100% divergent in the AP line (T base- same as reference) and AR line (C base). This gene is a member of the POU family, which are transcriptional regulators and help control cell type specific differentiation pathways. This gene has been identified as a tumor suppressor involved in Wilm Tumor (WT) hereditary predisposition (RefSeq, 2009; Perotti et al., 2004). Since this gene has already been identified as a tumor suppressor gene in humans, it is possible that it has the same function in chickens since the genomes are similar. This SNP is also located in an intron, but as mentioned with FGF 14, an introgenic mutation can affect gene expression.

Lastly, NCK2 (NCK adaptor protein 2), has a 100% SNP (A base) frequency for AR line and a 100% SNP (T base) frequency for AP line with AP matching the reference base. This gene is also located on an intron. NCK2 is involved in cell signaling, proliferation, cytoskeleton organization, and stress response. This gene is known to promote human melanoma cell proliferation, migration and invasion in melanoma-derived tumor growth. When injected with overexpressed human primary melanoma cells, mice experienced increased melanoma-derived tumor growth, indicating NCK2 can influence human melanoma phenotype (Labelle-Cote et al., 2011). Again, with the human and chicken genome containing 60% of corresponding genes, NCK2 could be a potential biomarker for tumor growth.

Phenotypic Results

Phenotypic differences in tumor development caused by RSV inoculation were depicted by a tumor score system (Figures 3-9). The results for the AR line, shown in Figure 10, confirm the line's ability to regress tumors over time compared to the tumor progression exhibited by the AP line (Figure 11). The AR line had a much quicker response to RSV than that of AP. On day 28, the average tumor score was 2.8 and then quickly dropped down to less than 1 by 40 days post inoculation (dpi). At the end of the trail, the average tumor score for AR was 1.125, which is about the size of a BB bullet. This phenotype was expected for the AR line. As shown in figure 11, the AP line does not show any signs of large tumors until day 28dpi. On this day, the average tumor score jumped to about 3, which is the size of a marble. On day 32dpi, the average tumor score drops to approximately 2, but then gradually increases to 4.25, the size of a ping pong ball. The gradual increase over time was expected for the AP line as they were selected for tumor progression. The difference between the average tumor scores at the end of the trial was significantly different, with a 95% Confidence Interval. The genotype-phenotype correlations for the AR and AP line were further tested in unrelated populations, discussed in Chapter 4.

CONCLUSION

One of the main objectives of this study was to link the phenotype of tumor progression and regression to specific genotypes, in this case SNPs. Whole genome sequencing was performed to identify SNPs that contribute to RSV induced tumors. Nine candidate SNPs, located in intron and exon regions, were identified to be potential biomarkers associated with RSV induced tumors. Sequencing results showed high divergence in point mutations between the AP and AR lines. The next step is to validate the candidate SNPs in unrelated populations. Identifying functional variants in the genome can be used in marker-assisted selection for performance improvement in farm livestock. Associations with functional variants can also be utilized for disease biology and pharmacology.

REFERENCES

Abel, Britain, Cara Willoughby, Sungchan Jang, Laura Cooper, Leike Xie, Chi Vo-Ransdell, and Grzegorz Sowa. (2012). N-terminal Tyrosine Phosphorylation of Caveolin-2 Negates Anti-proliferative Effect of Transforming Growth Factor Beta in Endothelial Cells. *FEBS Letters*. 3317-323.

Alfano, Massimo, and Guido Poli. (2005). Role of Cytokines and Chemokines in the Regulation of Innate Immunity and HIV Infection. *Molecular Immunology*. 161-82.

Ansari, Daniel, Roland Andersson, Monika P. Bauden, Bodil Andersson, Joanne B. Connolly, Charlotte Welinder, Agata Sasor, and György Marko-Varga. (2014). Protein Deep Sequencing Applied to Biobank Samples from Patients with Pancreatic Cancer. *Journal of Cancer Research and Clinical Oncology*. 369-80.

Bell, Daphne W., Nilabja Sikdar, Kyoo-Young Lee, Jessica C. Price, Raghunath Chatterjee, Hee-Dong Park, Jennifer Fox, Masamichi Ishiai, Meghan L. Rudd, Lana M. Pollock, Sarah K. Fogoros, Hassan Mohamed, Christin L. Hanigan, Suiyuan Zhang, Pedro Cruz, Gabriel Renaud, Nancy F. Hansen, Praveen F. Cherukuri, Bhavesh Borate, Kirk J. Mcmanus, Jan Stoepel, Payal Sipahimalani, Andrew K. Godwin, Dennis C. Sgroi, Maria J. Merino, Gene Elliot, Abdel Elkahloun, Charles Vinson, Minoru Takata, James C. Mullikin, Tyra G. Wolfsberg, Philip Hieter, Dae-Sik Lim, and Kyungjae Myung. . (2011). Predisposition to Cancer Caused by Genetic and Functional Defects of Mammalian Atad5. *PLos Genet*. e1002245.

Cooper, Geoffrey M., and Robert E. Hausman. ;, 2007. The Cell: A Molecular Approach. 4th ed. Washington, D.C.: ASM.

Fareed, Mohd, and Mohammad Afzal. (2012). Single Nucleotide Polymorphism in Genome-wide Association of Human Population: A Tool for Broad Spectrum Service. *Egyptian Journal of Medical Human Genetics.* 123-34.

Germer, S. (2000). High-Throughput SNP Allele-Frequency Determination in Pooled DNA Samples by Kinetic PCR. *Genome Research*. 258-66.

Goldfarb, M. (2001). Signaling By Fibroblast Growth Factors: The Inside Story. *Science Signaling*.

Goldfarb, M. (2001). Signaling By Fibroblast Growth Factors: The Inside Story.*Science's STKE : Signal Transduction Knowledge Environment*. 106, pe37.

Harland, M., S. Mistry, D. T. Bishop, and J. A. Newton Bishop. (2001).A Deep Intronic Mutation in CDKN2A Is Associated with Disease in a Subset of Melanoma Pedigrees. *Human Molecular Genetics*. 2679-686.

Huang, H.y., Z. Liang, S.f. Li, C.m. Li, Z.h. Zhao, and Q.b. Wang. (2015). Polymorphism Identification in BMP15 and GDF9 Genes and Their Association with Egg Production in Chickens*. British Poultry Science*. 277-83.

Kim, Hye-Eun, Dae-Ghon Kim, Kyung Jin Lee, Jang Geun Son, Min-Young Song, Young-Mi Park, Jae-Jung Kim, Sung-Won Cho, Sung-Gil Chi, Hyun Sub Cheong, Hyoung Doo Shin, Sang-Wook Lee, and Jong-Keuk Lee. (2012). Frequent Amplification of CENPF, GMNN and CDK13 Genes in Hepatocellular Carcinomas. *PLoS ONE*.

Kuk, A. Y. C., J. Xu, and Y. Yang. (2010). A Study of the Efficiency of Pooling in Haplotype Estimation. *Bioinformatics*. 2556-563.

Lee, Mi, Ence Yang, Mireille Morisson, Alain Vignal, Yong-Zhen Huang, Hans H Cheng, William M Muir, Susan J Lamont, Hyun Lillehoj, Sung Lee, and James E Womack. 2014). Mapping and Genotypic Analysis of the NK-lysin Gene in Chicken. *Genetics Selection Evolution*.43.

Lian, Ling, Daixi Zhang, Qiong Wang, Ning Yang, and Lujiang Qu. (2015). The Inhibitory Effects of Gga-miR-199-3p, Gga-miR-140-3p, and Gga-miR-221-5p in Marek's Disease Tumorigenesis. *Poultry Science*. 2131-135.

Luo, Chenglong, Hao Qu, Jie Ma, Jie Wang, Xiaoxiang Hu, Ning Li, and Dingming Shu. (2013). A Genome-wide Association Study Identifies Major Loci Affecting the Immune Response against Infectious Bronchitis Virus in Chicken. *Infection, Genetics and Evolution.* 351-58.

Matsuno, Akira, Haruko Katayama, Mineko Murakami, and Tadashi Nagashima. (2004). In Vivo Growth Suppression of Rat C6 Glioma Transplanted in Rat Brain Using Antisense Oligonucleotide for Microtubule-associated Protein 1A Messenger Ribonucleic Acid. British Journal of Neurosurgery Br J Neurosurg. 343-46.

Meijer, Azzi L., and J. Y. Wang. (1991). Cyclin B Targets P34cdc2 for Tyrosine Phosphorylation. *The EMBO Journal.* 75.4. 588-97.

Méndez-Lagares, Gema, Didiana Jaramillo-Ruiz, Marjorie Pion, Manuel Leal, M. A. Muñoz-Fernández, Yolanda M. Pacheco, and Rafael Correa-Rocha. (2014). HIV Infection Deregulates the Balance Between Regulatory T Cells and IL-2–Producing CD4 T Cells by Decreasing the Expression of the IL-2 Receptor in Treg. *AIDS Journal of Acquired Immune Deficiency Syndrome*. 278-82.

Meyer, Michael A. (2014). Highly Expressed Genes in Human High Grade Gliomas: Immunohistochemical Analysis of Data from the Human Protein Atlas. *Neurology International*

Min, W. (2004). Identification and Characterization of Chicken Interleukin-16 CDNA. *Developmental & Comparative Immunology.* 153-62.

Perotti, Daniela, Giovanna De Vecchi, Maria A. Testi, Elena Lualdi, Piergiorgio Modena, Patrizia Mondini, Fernando Ravagnani, Paola Collini, Francesca Di Renzo, Filippo Spreafico, Monica Terenziani, Gabriella Sozzi, Franca Fossati-Bellani, and Paolo Radice. (2004). Germline Mutations of ThePOU6F2 Gene in Wilms Tumors with Loss of Heterozygosity on Chromosome 7p14. *Human Mutation*. 400-07.

Sham, Pak, Joel S. Bader, Ian Craig, Michael O'donovan, and Michael Owen. (2002). DNA Pooling: A Tool for Large-scale Association Studies. *Nature Reviews Genetics*. 862-71.

Soengas, M. S. (1999). Apaf-1 and Caspase-9 in P53-Dependent Apoptosis and Tumor Inhibition. *Science*. 156-59.

Untergasser, A., H. Nijveen, X. Rao, T. Bisseling, R. Geurts, and J. A. M. Leunissen. 2007. Primer3Plus, an enhanced web interface to Primer3. http://www.bioinformatics.nl/cgibin/primer3plus/primer3plus.cgi/ Accessed 9/10, 2015.

Yang, Y., J. Zhang, J. Hoh, F. Matsuda, P. Xu, M. Lathrop, and J. Ott. (2003). Efficiency of Single-nucleotide Polymorphism Haplotype Estimation from Pooled DNA. *Proceedings of the National Academy of Sciences.* 7225-230.

Yellapa, Aparna, Janice M. Bahr, Pincas Bitterman, Jacques S. Abramowicz, Seby L. Edassery, Krishna Penumatsa, Sanjib Basu, Jacob Rotmensch, and Animesh Barua. (2012). Association of Interleukin 16 With the Development of Ovarian Tumor and Tumor-Associated Neoangiogenesis in Laying Hen Model of Spontaneous Ovarian Cancer. *International Journal of Gynecological Cancer.* 199-207.

LEGEND OF TABLES AND FIGURES

Table 1: Primer sets for 9 candidate biomarkers

Candidate SNPs had a forward, reverse, and sequencing primer. This table includes the SNP location, GC% content, annealing temperature and product size for each candidate biomarker.

Table 2: Tumor Sizing Scale

Table 3: Sequencing Results for larger AP and AR populations

SNP frequencies for each candidate are listed as well as the type of mutation, genotype, feature, and impact. SNP frequencies are listed as SNP % of called base in this table. The lowest SNP frequency of the candidate biomarkers was MAP1A at 60% SNP frequency.

Figure 1: PCR 96 well plate format for AP and AR lines

Top four rows contained AR samples and bottom four rows contained AP samples. A 96-well plate with this set up was used for each primer set. The order of DNA samples remained the same for each PCR.

Figure 2: Sample pooling for PCR purification and sequencing

Four different primer sets and 8µl from each sample pooled into one 96-well plate totaling 32µl of 4 different PCR products. After PCR purification, each well contained 100µl of pooled DNA. Then 3µl of all samples in each row were pooled into one well. For example, 3µl of all samples in row A are pooled into one tube.

Figures 3-9: Photos of each tumor score shown as examples for each score 1-6

Figure 10: Average tumor scores for AP and AR chickens with Standard Error

Average tumors scores over time are shown. The y-axis represents the tumor score and the x-axis represents the number of days post inoculation. It is shown that the average tumor scores increased over time and did not regress for the AP chickens. It is shown that the average tumor scores increased at about 28 days post inoculation, but then decreased shortly after and remaining the same for the AR chickens.

Table 1: Primer sets for 9 candidate biomarkers

SNP location (Chr:bp)	Feature name	Primer name	Primer sequences	GC%	T_m (°C)	Product size (bp)
1:134690198	NCK2	NCK2_1_F	caggtccttagctgctttctgt	50	60.07	471
		NCK2_1_R	acaggaaagctgatgggattta	40.91	59.97	
		NCK2_Seq_F_1	actccagcaagctgttgatcct	50	54.8	
2:50117844	POUF6	POUF6_1_F	gacctctgtctctgaaagtagc	50	54.8	485
		POUF6_1_R	tttggcatcagactccaacacc	50	54.8	
		POUF6_Seq_1_F	gcttctacacattctgtcttgg	45	53	
1:3380404	IL2RA	IL2RA_1_F	tcctgaaccccaagcaggtttt	50	54.8	476
		IL2RA_1_R	acaaagcagcacgttgttaggc	50	54.8	
		IL2RA_1_SEQ_F	cttaggttgttcctgtgttggc	50	54.8	
18:6611691	ATAD5	ATAD5_1_F	ggttttcctgaaccttgctacg	50	54.8	462
		ATAD5_1_R	cacgtactggacaagcaccaaa	50	54.8	
		ATAD5_1_SEQ_F	caaggtggttcaagaacttcag	45	53	
21:4919322	CASP9	CASP9_1_F	ttctcgttccctccttctgtct	50	54.8	434
		CASP9_1_R	aaaaccaggacttcgcagccta	50	54.8	
		CASP9_Seq_F	agttggttgttgccaacactgc	50	54.8	
1:142233297	FGF14	FGF14_F_1	gttaacacctgcaggctcacca	55	56.7	520
		FGF14_R_1	gtacaggtcgtacaatccttac	45	53	
		FGF14_Seq_1	aatggcacttgtgtctgcctg	50	54.8	
2:50469650	CDK13	CDK13_F_1	gggcactgcctttcctattttg	50	54.8	448
		CDK13_R_1	tgtactgctcgtaggagaaggt	50	54.8	
		CDK13_Seq_1	gaaacaggcgatcagcaaaatc	45	53	
10:11603087	IL16	IL16_F_1	tggttgcttcctgctcttgatg	50	54.8	510
		IL16_R_1	acacttaacgcgtctcagcatt	45	53	
		IL16_Seq_1	ggaaacaggacacttctgctaa	45	53	
10:19701184	MAP1A	MAP1A_F_1	aggagatgccatcttgactgtg	50	54.8	451
		MAP1A_R_1	agtgaacagcatgccttcctct	50	54.8	
		MAP1A_Seq_1	tcagcctctgcagagtcaacct	55	56.7	

Table 2: Tumor Sizing Scale

Tumor Score	Size Range (cm^2)	Subjective Score
0	0	No tumor
1	0.1 – 0.25	BB
2	0.25 – 1.0	Pea
3	1.0 – 4.0	Marble
4	4.0 – 9.0	Ping pong ball
5	9.0 – 16.0	Golf ball
6	16.0 and greater	Egg

Table 3: Sequencing Results for larger AP and AR populations

Gene	Line	Chr	Ref Position	Ref Base	Mutation	Called Base	Expected Base	SNP % of called base	Genotype	Feature	Impact
IL16	AP	10	11603087	T	Non-synonymous	C	C	100% C	Homo. Var	CDS	Ile to Thr (ATT -> ACT)
FGF14	AP	1	142233298	T	Non-coding region	C	C	100% C	Homo. Var	gene	
IL2RA	AP	1	3380404	C	Silent	T	T	100% T	Homo. Var	CDS	Ile to Ile (ATC->ATT)
MAP1A	AR	10	19701184	G	Non-synonymous	A:G	A	60% A	Hetero. Var	CDS	Gly to Arg (GGA->AGA)
CDK13	AR	2	50496950	G	Non-synonymous	A	A	100% A	Homo. Var	CDS	Thrto Thr (ACG->ACA)
ATAD5	AR	18	6611691	C	Stop codon	T:C	T	83% T	Hetero. Var	CDS	Gln to stop (CAG->TAG)
CASP9	AR	21	4919322	G	Nonsense	A:T	A	83% A	Hetero. Var	CDS	Cys to stop (TGC->TAG)
POU6F2	AR	2	50117844	T	Non-coding region	C	C	100%	Homo. Var	gene	
NCK2	AR	1	134690198	T	Non-coding region	A	A	100%	Homo. Var	gene	

Figure 1: 96-format for AP and AR lines

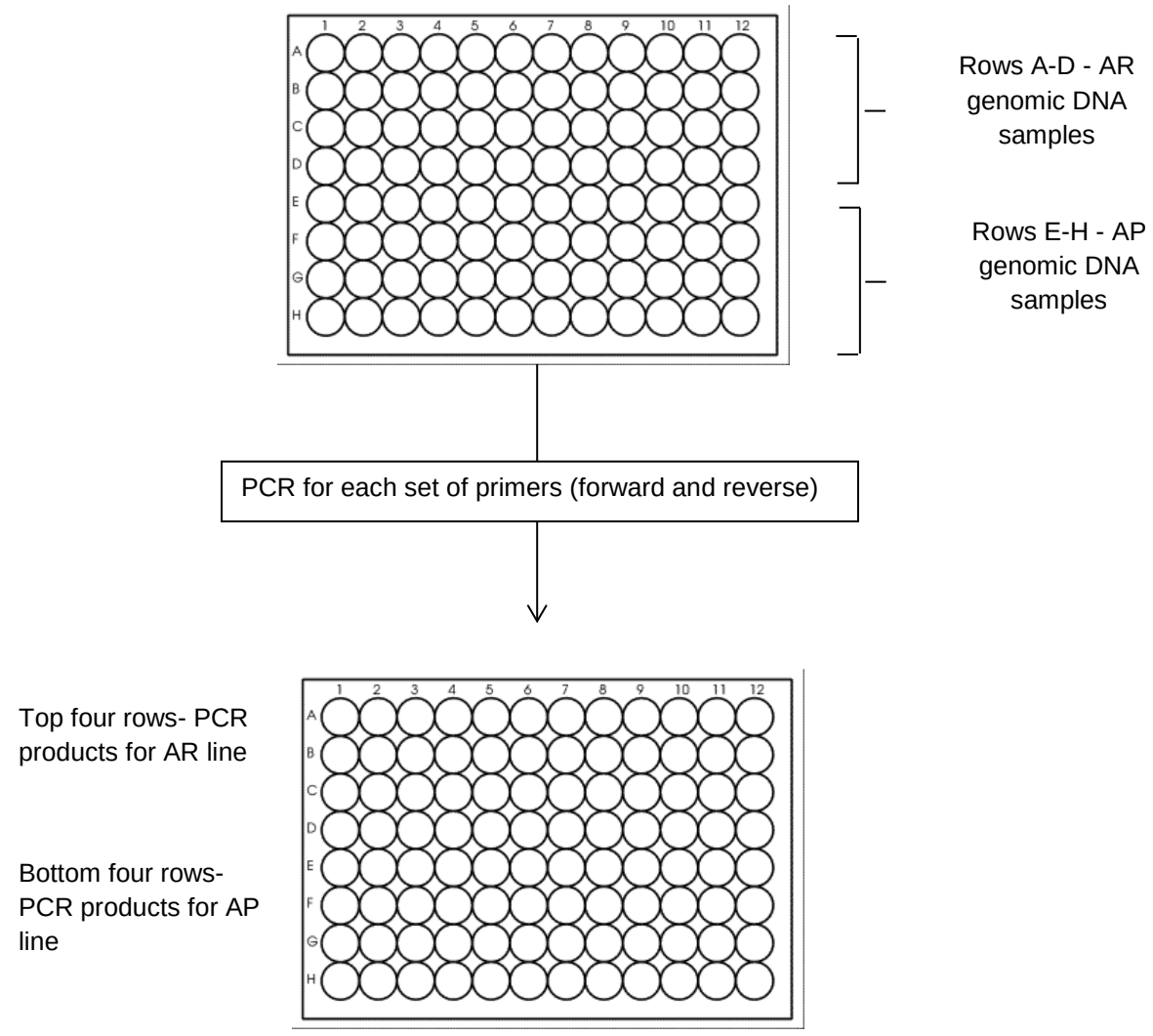


Figure 2: Sample pooling for PCR purification and sequencing.

***** Note: each well represents the same sample/same bird in each plate.

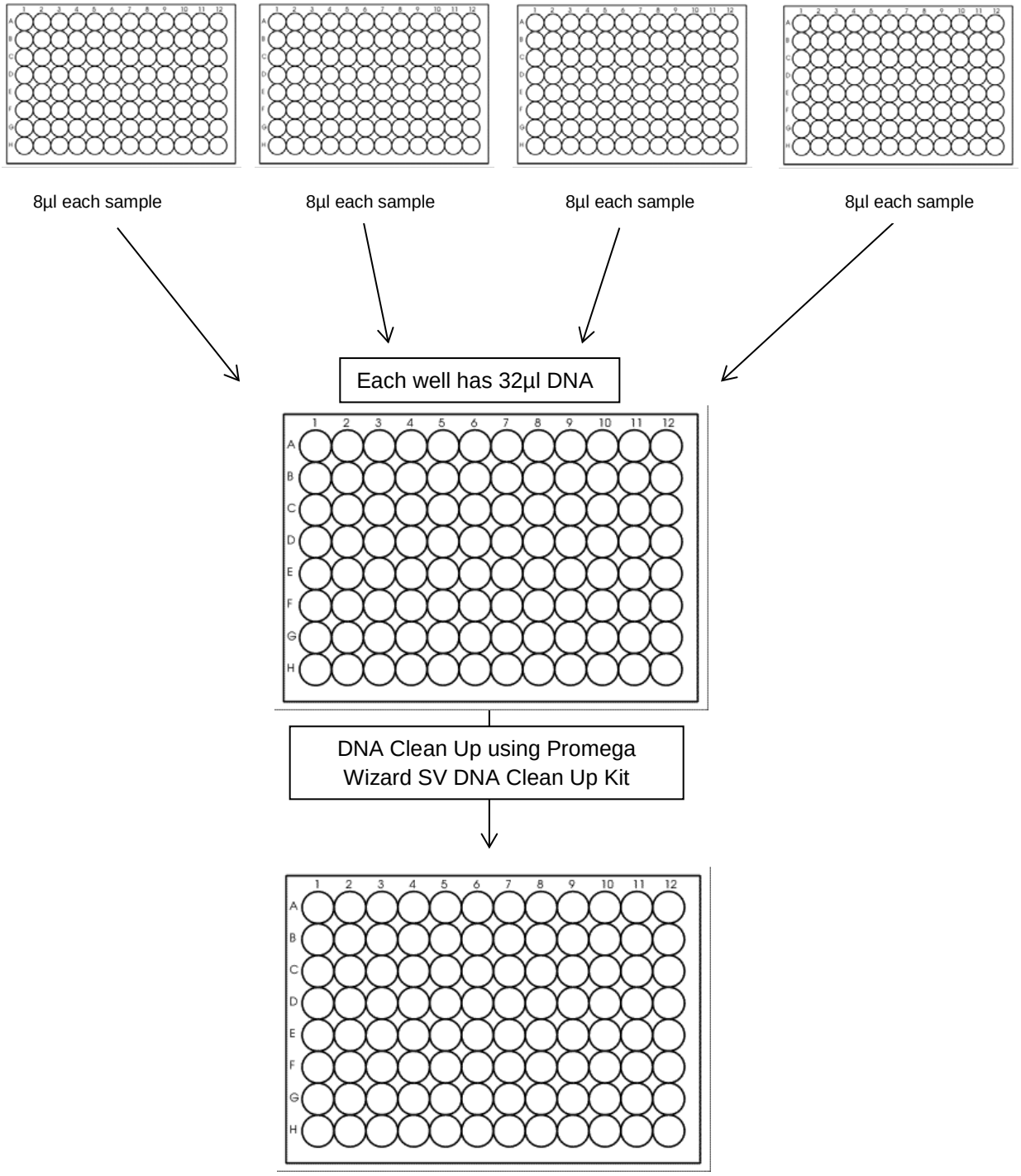
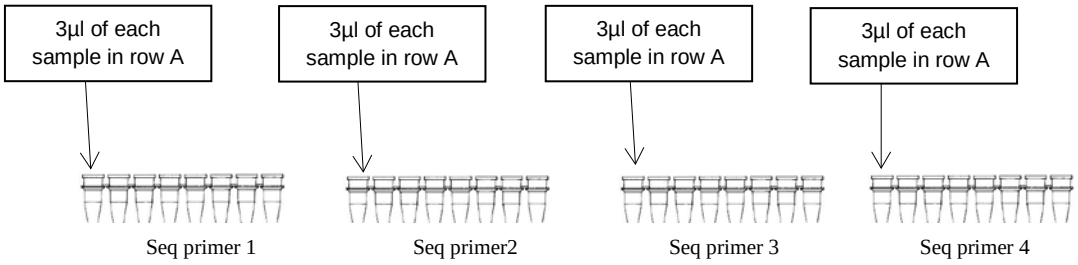


Figure 2 cont'd: Each tube displayed below represents a row with a specific sequencing primer



Each tube contains 1.5µl sequencing primer, 20ng DNA, and distilled water to a total volume of 13µl. Each set of 8 tubes contain the same DNA but have different primers in order to sequence different PCR products in the pooled samples. Samples are then submitted for sequencing to the DNA Core Lab.

Figure 3: Wing web without tumor

Figure 4: Tumor with score of 1 (BB size)

Figure 5: Tumor with score of 2 (pea size)

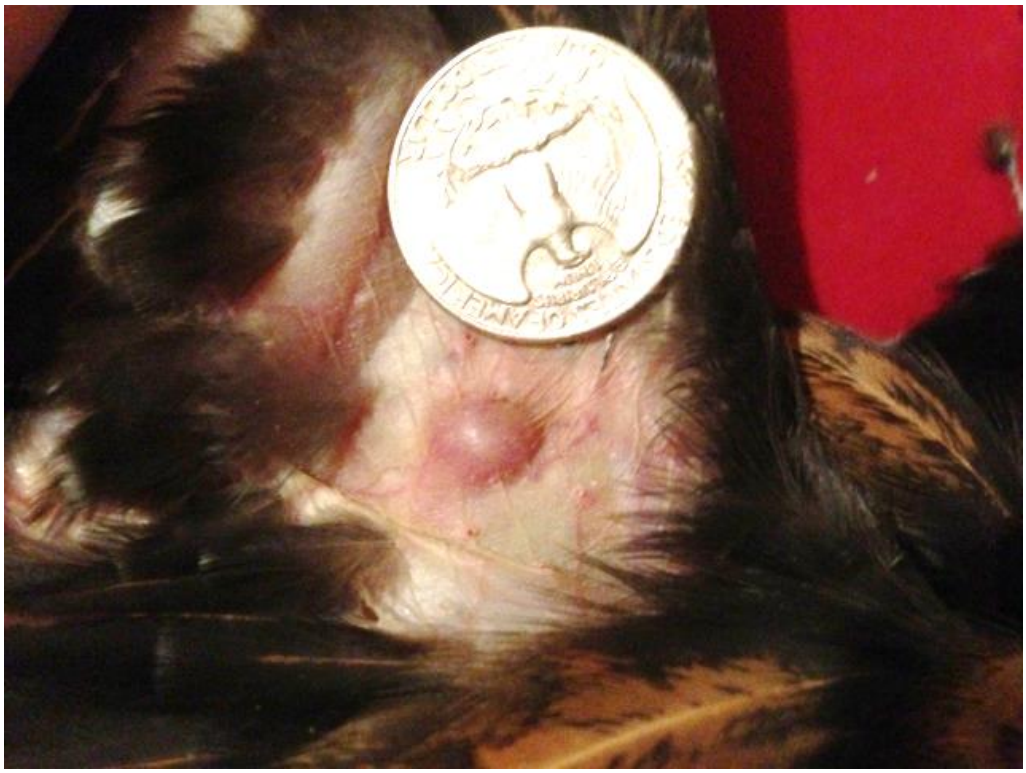


Figure 6: Tumor with score of 3 (marble size)

Figure 7: Tumor with score of 4 (ping pong ball size)

Figure 8: Tumor with score of 5 (golf ball size)

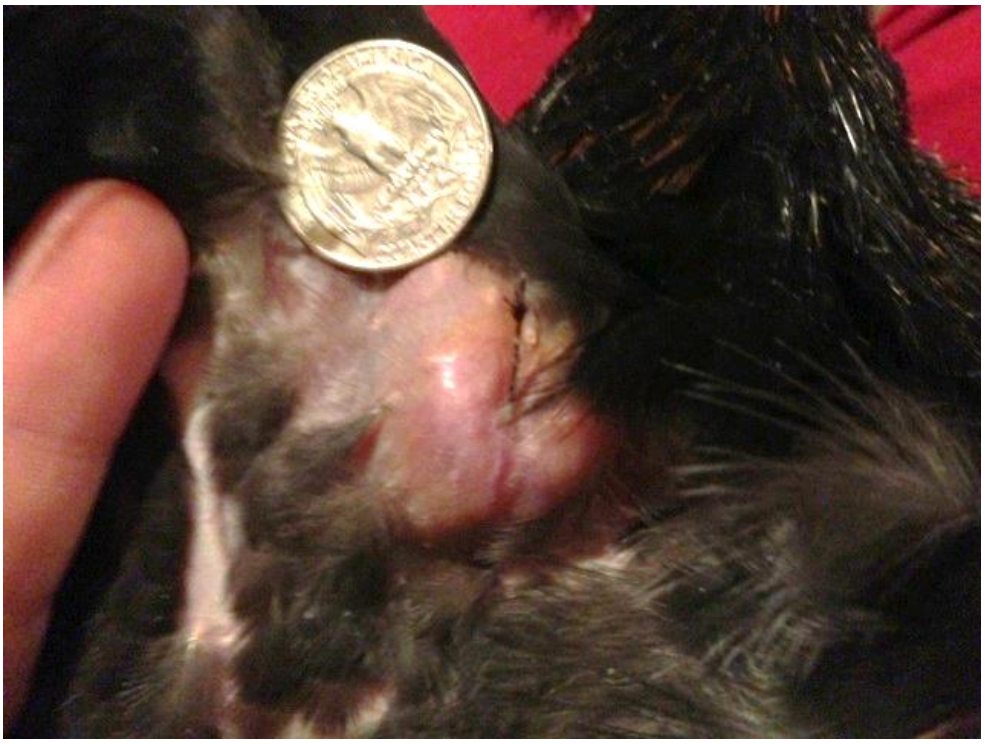


Figure 9: Tumors with score of 6 (anything larger than a golf ball and has coloration and/or fluid oozing)

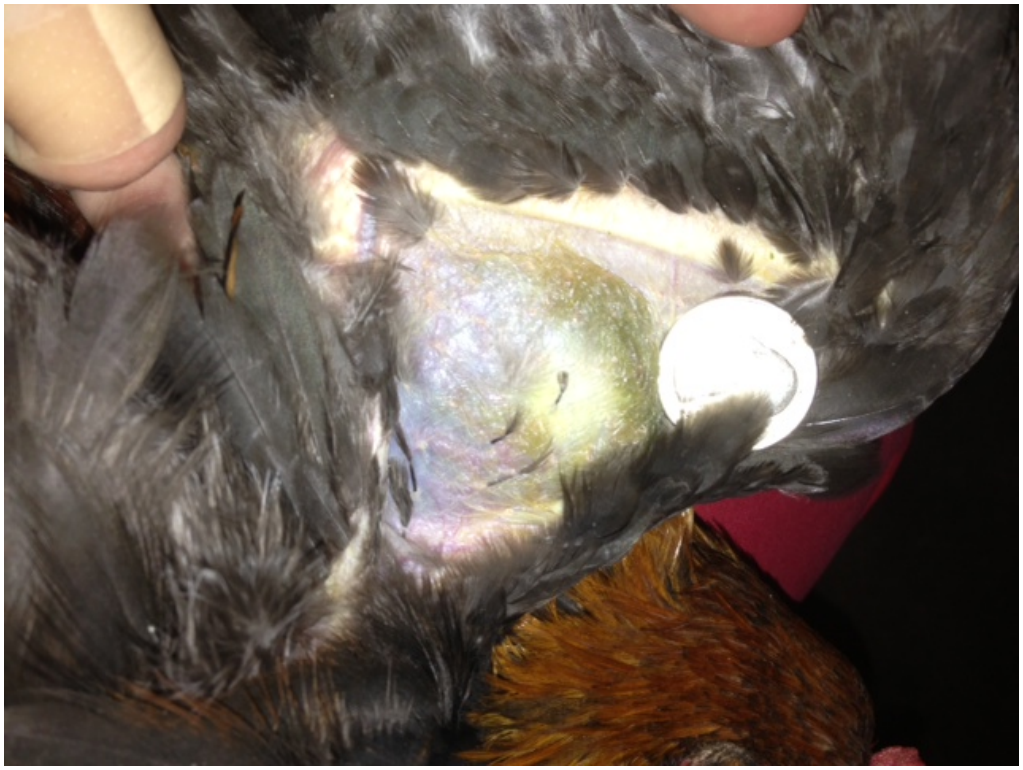
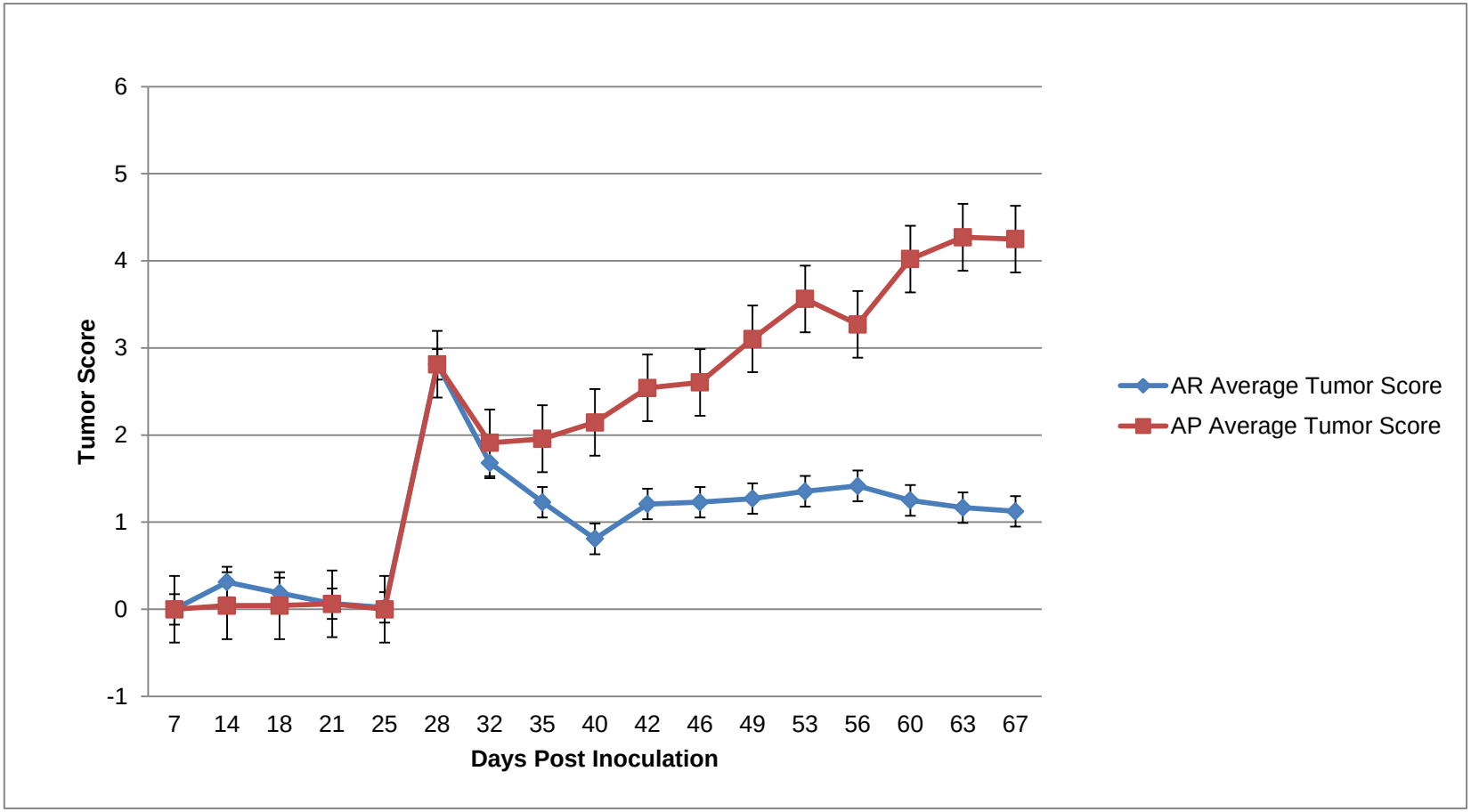


Figure 10: Average tumor scores for AP and AR chickens



CHAPTER 4: APPLICATION OF CANDIDATE GENETIC MARKERS IN GIANT JUNGLE FOWL AND WHITE LEGHORN POPULATIONS

INTRODUCTION

Studies to identify genetic markers for regression and progression from RSV-induced tumors have been performed in the past (Spanakos, 2007). Many results point to the MHC region (B-complex) of having the most indicative loci for tumor regression and progression. Spanakos (2007) found four loci within the B-complex associated with tumor response induced by RSV; the Ea-A, Ea-C, Ea-I and Y loci. Different haplotpyes were investigated and the $A^{4}A^{5}B^{13}B^{13}C^{2}$ $C^{2}I^{2}I^{8}Y^{1.1}Y^{7}$ haplotype was determined to be the most progressive of birds with the $B^{13}B^{13}$ genotype. A structural analysis of MHC alleles in RSV tumor regression chickens using a BAC library uncovered great genetic diversity. The BLB2 gene, encoding the MHC class II beta chain, had 19 non-synonymous SNPs; however, a comparison of WLA (RSV tumor regressive allele) and B6 haplotypes associated with tumor regression revealed that the BLB2 gene was not indicative of tumor regression intensity (Suzuki et.al., 2012). While the MHC region in the chicken genome plays an important role in the immune response to RSV, there is evidence showing non-MHC loci are also involved. Birds with the MHC genotypes found to be susceptible to RSV have been observed to regress tumors, while some birds with resistant genotypes succumb to RSV. Certain combinations of haplotypes could be the underlying cause of regression or progression of RSV-induced tumors.

In this study, nine different SNPs were identified as being potential biomarkers associated with RSV induced tumors in chickens and then validated for degree of divergence in the AR and AP lines. The high SNP frequency differences between the two lines showed significant differences, indicating the need to test the genotype-phenotype correlation. The Giant Jungle Fowl (GJF) and White Leghorn (WL) lines were chosen as the unrelated populations to validate the association of the potential genetic markers to tumor growth after RSV infection. Candidate SNPs were validated in each line to correlate the phenotypes and genotypes to tumors developed from RSV. Genotypic and phenotypic data was sampled and analyzed to find such correlations. In this study, it was hypothesized that the GJF population would have a higher percentage of regression over progression since the AR line was derived from this breed. The WL population was hypothesized to have a higher percentage of progression, but not higher than the AP line established at the University of Arkansas. The AP line was derived from a cross between GJF and

WL chickens, thus the WL line is more similar to the AP line genetically. In this chapter, the nine potential SNP biomarkers identified by genome sequencing and validation studies of the AP and AR chicken lines were examined to determine whether the chosen SNPs are associated with RSV-induced tumor progression and regression in the two unrelated populations GJF and WL.

MATERIALS AND METHODS

Genetic Lines

Two unrelated populations of chicken lines, GFJ and WL, were utilized to determine SNP frequencies in this study. The GJF line is theorized to be highly inbred as its origins stem from one male and five hens from Southeast Asia in 1951 (Gyles et al., 1966). This strain has been maintained at the AES since 1951 as a small closed flock. The original WL strain used to create the AP line is no longer being maintained at the AES, thus a commercial WL population from Kimbrough Farms in Neosho, MO was used as the other unrelated chicken population.

DNA sampling and preparation

Whole blood was collected from the GJF and WL lines via wing web puncture from more than 100 birds at two weeks of age. Approximately 2-3 drops (100-200µL) of blood was collected and stored in tubes containing 100µl sodium citrate ($C_6H_8O_7$) to prevent clotting. Genomic DNA was extracted into a 96 well format using the Wizard SV 96 Genomic DNA Purification System (Promega; Madison, WI) following manufacturer's instructions with some modifications. Whole blood (10µL) was thoroughly mixed with proteinase K (6.7µL/µL) and incubated at room temperature for 10 minutes. The lysis buffer (300µL) was and triton X-300 (33µL) was missed with RNase A (55µL) in a single boat. This mixture was added to the blood/proteinase K mixture and blood clots were dissociated by repetitive pipetting. Once blood clots were dissociated, a vacuum manifold system was used to filter lysate through filters of the binding plate. Three consecutive cycles of ethanol washing was applied followed by elution with nuclease- free water (200µL). Extracted DNA concentration was quantified using NanoDrop 1000 spectrophotometer (Thermo

Fisher Scientific Inc.; Waltham, MA). For all samples, a 1ng/µL dilution was prepared in empty 96 well plates for PCR and used as an aliquot.

PCR

For GJF and WL lines, tumor scores were analyzed and birds with more extreme tumor progression or regression were used for SNP validation in the unrelated populations. The GJF line had 50 out of 96 samples representing tumor regression, or tumor resistance (RES) and 34 out of 96 samples representing tumor progression, or tumor susceptibility (SUS). The 50 RES and 34 SUS samples were pooled to create two pooled samples representing both phenotypes. The WL line had 27 RES samples and 21 SUS samples, which were pooled to represent one sample for each phenotype. PCR was performed for each of the nine candidate genes for GJF and WL lines, resulting in 18 different PCR reactions for each line. The same primer sets used in Chapter 3 were used for GJF and WL PCR reactions. Total reaction volume was 25µL: 5µL DNA (1ng/µL), 2.5µL 10x buffer, 1µL 2.5mM dNTP mix, 2µL F primer, 2µL R primer, 0.5µL Taq DNA Polymerase (NEB, Ipswich, MA). PCR conditions were as follows: denaturation at 95ºC for 1 min, 40 cycles of amplification (95ºC for 30 sec, 60ºC for 1 min, 72ºC for 1 min), and a final extension at 72ºC for 10 min. PCR assays were performed using Applied Biosystems 2720 Thermal Cycler (Life Technologies, Carlsbad, CA). The quality and validation of PCR product was performed by 1% agarose gel electrophoresis.

Gel Electrophoresis

From each PCR product, 5µL was mixed with 6x loading dye, which contained bromophenol blue and xylene cyanol dyes. The mixtures were loaded into wells of a 1% agarose gel prepared with 1x TAE buffer solution. The first well was loaded with 5µL Hi-Lo DNA molecular weight marker (Bionexus, Inc., Oakland, CA) to confirm the PCR product size. A Bio-Rad Gel Doc XR imaging system (Bio-Rad Laboratories, Hercules, CA) was used for gel imaging.

PCR product purification

After gel electrophoresis, PCR product was purified using the Illustra GFX PCR DNA purification kit (GE Healthcare Life Sciences, Pittsburg, PA) following manufacturer's protocol. Membrane-binding solution (100µL) was added to the pooled PCR products were centrifuged at 10,000X. Washing buffer was applied to the samples and then eluted with 30µL nuclease-free water. The concentrations of purified DNA were quantified using NanoDrop 1000.

Sanger Sequencing

To prepare the purified PCR products for sequencing, pooled DNA with concentration 10ng/µl, diluted with distilled water, was mixed with 4-5µl of the sequencing primers with concentration 1.6uM. Sequencing was performed at the University of Arkansas for Medical Sciences (UAMS) in Little Rock, Arkansas. Sequencing results were analyzed with Sequence Scanner software v.1 (Applied Biosystems of Thermo Fisher, Waltham, MA). Peak heights at SNP locations were examined and the ratios of different bases at the SNP location were recorded and analyzed.

RSV Inoculation and blood collection

The bird populations being tested were also observed for phenotypic data collection. Approximately 100 birds from each line were hatched and maintained at the AES (Fayetteville, Arkansas). To produce the GJF birds for this study, matings were achieved via artificial insemination using pooled semen of at least 10 males or more. At hatch, all chicks were given a wing band with a unique identifying number and then placed randomly in pens, where feed and water were provided *ad libitum* during grow-out of the birds. Tumor scores were recorded for each bird/wing band about every 3 days.

All birds were inoculated subcutaneously with 70pfu Bryan high-titer strain Rous Sarcoma Virus (0.1mL) in the wing web at six weeks of age. RSV-induced tumors were examined three times a week and recorded for 8-10 weeks post inoculation to tumor progression and regression. The same sizing scale described in Chapter 3 was used for recording tumor scores. Images of example tumor sizes can be reviewed in Chapter 3, Figures 3-9.

Once a tumor reached a score of 5 for one week, and showed no sign of regression, the bird was humanely euthanized. Also, if the skin covering the tumor broke open, and the contents were visible outside the tumor, the bird was humanely euthanized. All birds were humanely euthanized 10 weeks post inoculation by CO_2 (Guidelines for Euthanasia for Poultry, American Veterinary Medical Association).

RESULTS AND DISCUSSION

The GJF and WL lines were used to determine whether the candidate SNPs would diverge relative to tumor incidence. Roughly 100 birds from each line were inoculated with RSV. In the GJF and WL lines, the most tumor resistant (RES) and tumor susceptible (SUS) birds were separated by phenotype and then pooled to determine approximate allele frequencies before testing individual SNPs. RES was classified as having a tumor reach a score of at least 3 or more and then regress to either a 1 or 0 for the last 3 consecutive scorings. SUS was classified as having a tumor develop and reach a score of at least 4 and remain at that score or higher. The GJF line had 50 RES samples pooled and 34 SUS samples pooled, totaling 84 samples. GJF RES and SUS birds represented 52% and 35%, respectively, of the entire GJF population tested (96 samples). The WL line had 27 RES samples pooled and 21 SUS samples pooled, representing 28% and 21%, respectively, of the entire WL population (96 samples). Overall, the sequencing results showed no significant differences at the SNP locations between the resistant (RES) and susceptible (SUS) birds within GJF and WL lines; however, there were SNPs which were divergent between the GJF and WL lines (Table 1).

The GJF line was derived from an unselected Giant Jungle Fowl population in Southeast Asia, the natural range of the Red Jungle Fowl (RJF) (Storey et al., 2012). Using DNA fingerprinting, GJF, also known as the Domestic Jungle Fowl, has been theorized to be more closely related to RJF than White Leghorn strains (Seigel, P.B. et al., 1992), thus its genome should be genetically more similar to the RJF genome than that of the WL line. GJF and WL showed means 33% and 60% differences to RJF for all tested SNPs; respectively, thus the WL line was expected to be more susceptible to RSV-induced tumor development since the GJF line is more regressive and more closely related to the RJF. The grand mean tumor score for GJF was 1.44 with a variance of 0.48 and the grand mean tumor score for WL was 2.92

with a variance of 0.33. An unpaired t-test showed a two-tailed p value of less than 0.00001 with a significance level (p-value) of 0.05, indicating a significant difference between the grand mean tumor scores for GJF and WL. The divergent SNPs between GJF and WL lines could be evidence of human-driven selection based on a phenotypic trait. For IL2RA, WL had a 100% SNP frequency (T base), differing from GJF and RJF (C base). There was also a 100% SNP frequency (A base) seen for WL NCK2, while GJF had a SNP frequency (A base) of 14% and 9% for resistant and susceptible birds, respectively. In the GJF line, SNPs in genes CASP9 and IL16 had no genetic difference to RJF and a 50% difference in SNPs housed in POU6F2, MAP1A, FGF14, and ATAD1. The WL line showed no genetic difference to RJF for FGF14 and MAP1A and 100% difference to RJF in POU6F2 and ATAD1. SNP frequencies for WL CDK13, IL16, CASP 9, and ATAD1 were 50%, 20%, 67%, and 92% respectively (Table 1).

Potential correlations associated with tumor progression and regression based on the chosen SNPs could not be analyzed since there were no genotypic differences between birds of differing phenotypes within the GJF or WL lines. All RES and SUS birds within each population had the same SNP frequencies. There were two SNPs which did show some differences between the tumor RES and SUS birds in the GJF lines. GJF NCK2 had a 5% difference between RES and SUS birds and GJF CDK13 had a 6% difference between RES and SUS birds. Chi-square analysis was performed and results showed no significant difference in either SNP ($p=0.3566$ for NCK2 and $p=0.2488$ for CDK13). A chi-square analysis could not be performed on any other SNP frequencies because all individuals within a population shared the same SNP frequency.

The GJF and WL lines represented two un-related populations to AP and AR and were not to be compared to each other. The genetic variation of the SNP locations in GJF and WL lines was not of interest in this study; however, it did not go unnoticed. There was no correlation between the phenotype and genotype of birds within a population. It is possible that during intense selection of AP and AR lines, genetic drift of the SNPs of interest became fixed, resulting in different genotypic combinations without causing phenotypic changes. Artificially selecting individuals could increase the chance of genetic drift, especially when only the phenotype is used as the selected trait. Limited gene flow and genetic drift in a

finite sub-divided population could potentially cause linkage disequilibrium, making it difficult to identify direct genotype-phenotype correlations (Ohta, 1982). The high frequency SNPs that are highly divergent from the reference do show evidence of genetic variation caused by selection. In order to find haplotype combinations related to tumor incidence caused by RSV, an analysis of variance could be performed for birds showing genetic differences within larger population.

In this chapter, nine candidate genetic markers were evaluated in GJF and WL populations, representing two populations un-related to the AP and AR lines. GJF and WL populations were challenged with RSV and then genotyped at the candidate SNP loci to identify any association to phenotypic change. No statistical significance was found between phenotypes within each line. The only SNP differences found were between the GJF and WL, which could be evidence of chicken domestication since the GJF is more closely related to RJF. The WL line has been selected for egg-laying performance and thus, more genetic drift caused by bottlenecking is likely. The GJF showed 33% difference from RJF at the nine SNP locations while WL showed 56% difference. To determine whether domestication is exhibited at the nine SNP locations, a larger sample size would need to be studied.

REFERENCES

Cordell, H. (2002). Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics,* 2463-2468.

Janeway C.A. Jr, Travers P., Walport M., Sclomchik, M.J. (2001). Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science.

Ohta, T. (1982). Linkage Disequilibrium due to Random Genetic Drift in Finite Subdivided Populations. *Proceedings of the National Academy of Sciences,* 5101-5101.

Siegel, P., Haberfeld, A., Mukherjee, T., Stallard, L., Marks, H., Anthony, N., & Dunnington, E. (1992). Jungle fowl-domestic fowl relationships: A use of DNA fingerprinting. *World's Poultry Science Journal Worlds Poult Sci J,* 147-155.

Shanmugasundaram, R.,and Selvaraj, R. (2011). Regulatory T Cell Properties of Chicken CD4 CD25 Cells. *The Journal of Immunology,* 1997-2002.

Spanakos, M. (2007). The effects of MHC and non-MHC genes on Rous sarcoma virus tumor development in the Arkansas progressor and regressor chicken lines. *ProQuest/UMI*.

Storey, A. A., Athens, J. S., Bryant, D., Carson, M., Emery, K., deFrance, S.,Higham, C., Huynen, L., Intoh, M., Jones, S.,Kirch, P. V., Ladefoged, T., McCoy, P., Morales-Muniz, A., Quiroz, D., Reitz, E. Robines, J., Walter, R., and Matisoo-Smith, E. (2012). Investigating the Global Dispersal of Chickens in Prehistory Using Ancient Mitochondrial DNA Signatures. *PLoS ONE*, *7*(7), e39171.

Suzuki, K., Matsumoto, T., Kobayashi, E., Uenishi, H., Churkina, I., Plastow, G., Yamashit, H., Hamasima, N., Mitsuhashi, T. (2010). Genotypes of chicken major histocompatibility complex B locus associated with regression of Rous sarcoma virus J-strain tumors. *Poultry Science,* 651-657.

Suzuki, K., Kobayashi, E., Yamashita, H., Uenishi, H., Churkina, I., Plastow, G., Hamasima, N., Mitsuhashi, T. (2011). Structural analysis of MHC alleles in an RSV tumour regression chicken using a BAC library. *Animal Genetics,* 348-351.

LEGEND OF TABLES AND FIGURES

Table 1: Frequencies for each candidate SNP for RES and SUS in GJF and WL lines

Observed SNP frequencies are listed for RES and SUS birds for both un-related chicken lines (GJF and WL). SNP frequency differences are listed between RES and SUS birds, between GJF/WL and RJF, and between GJF and WL

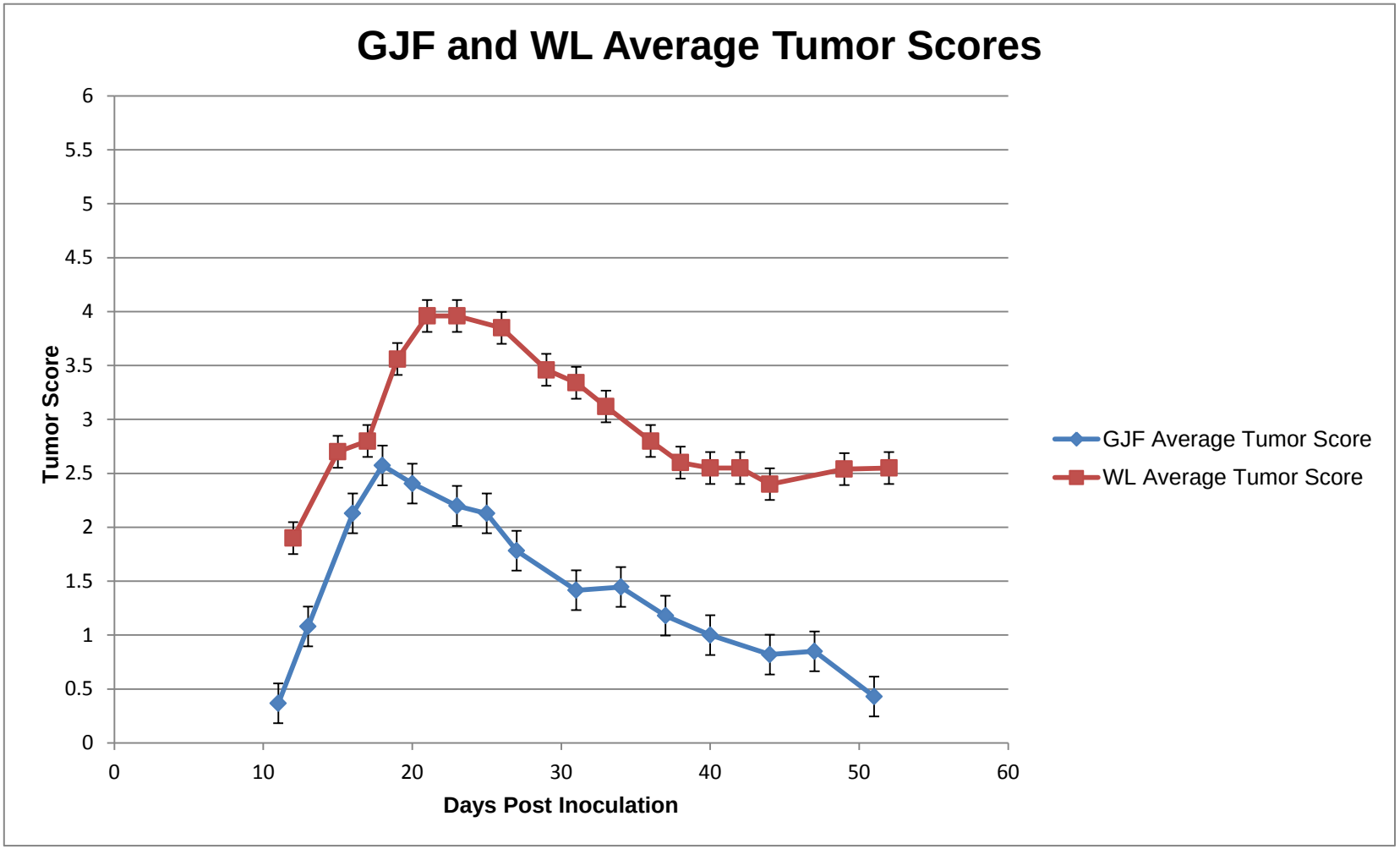
Figure 1: GJF and WL Average Tumor Scores with Standard Error

Average tumor scores for GJF and WL population (96 samples) with moving average trendline and standard error.

Table 1: Frequencies for each candidate SNP for RES and SUS in GJF and WL lines

Gene	Ref	Pheno	GJF	GJF ratio	SNP freq	% diff to ref	WL	WL ratio	SNP freq	% diff to ref	% diff b/w lines
ATAD1	C	RES	C/T	1:1	0.50	0.50	T/A	12:1	0.92	1.00	0.42
ATAD1	C	SUS	C/T	1:1	0.50	0.50	T/A	12:1	0.92	1.00	0.42
CASP9	G	RES	G	1:0	0.00	0.00	A/G	2:1	0.67	0.67	0.67
CASP9	G	SUS	G	1:0	0.00	0.00	A/G	2:1	0.67	0.67	0.67
CDK13	G	RES	A/G	5:1	0.83	0.83	A/G	1:1	0.50	0.5	0.33
CDK13	G	SUS	A/G	8:1	0.89	0.89	A/G	1:1	0.50	0.5	0.39
FGF14	T	RES	C/T	1:1	0.50	0.50	T	1:0	0.00	0.00	0.50
FGF14	T	SUS	C/T	1:1	0.50	0.50	T	1:0	0.00	0.00	0.50
IL2RA	C	RES	C	1:0	0.00	0.00	T	0:1	1.00	1.00	1.00
IL2RA	C	SUS	C	1:0	0.00	0.00	T	0:1	1.00	1.00	1.00
IL16	T	RES	T	1:0	0.00	0.00	T/C	4:1	0.20	0.20	0.20
IL16	T	SUS	T	1:0	0.00	0.00	T/C	4:1	0.20	0.20	0.20
MAP1A	G	RES	G/A	1:1	0.50	0.50	G	1:0	0.00	0.00	0.50
MAP1A	G	SUS	G/A	1:1	0.50	0.50	G	1:0	0.00	0.00	0.50
NCK2	T	RES	T/A	7:1	0.14	0.14	A	0:1	1.00	1.00	0.91
NCK2	T	SUS	T/A	10:1	0.09	0.09	A	0:1	1.00	1.00	0.91
POU6F2	T	RES	T/C	1:1	0.50	0.50	C	0:1	1.00	1.00	0.50
POU6F2	T	SUS	T/C	1:1	0.50	0.50	C	0:1	1.00	1.00	0.50

Figure 1: GJF and WL Average Tumor Scores with a 2 per. Mov. Avg



CONCLUSION

Poultry is a key organism in genetic research due to breeding feasibility, relatively short generation periods and distinct phenotypes. In addition, poultry experimental populations serve as models in base and applied biomedical studies for human disease. In addition, recent development of new sequencing technologies enables vertebrate genomes to be sequenced more efficiently. Hence, genome sequencing of commercial and experimental chicken lines can provide insight into the genetic correlates of phenotypic changes. Whole genome sequencing can improve the understanding of poultry diversity, provide a tool for the use of genome-wide marker assisted selection to enhance genetic gain in commercial stock, yield knowledge of genetic diagnostics to detect predisposition of genetic disorders, and can support characterization of potential biological mechanisms of genetic disease.

Genetic correlations between specific genotypes to phenotypic traits can increase heritability in a population. Increased heritability can enable rapid change under selection pressures for traits of interest. Genomic data can provide the poultry and biomedical industries biomarkers which can be used in animal breeding programs and prognosis of certain disease in human medicine. One type of biomarker is a single nucleotide polymorphism (SNP) and has been widely used for its ease of detection. Application of next generation sequencing to agricultural and biomedical chicken populations can identify SNPs associated with important phenotypic traits.

Genetic lines representing tumor progression (AP) and regression (AR) after inoculation of rous sarcoma virus (RSV) have been established and maintained by N. B. Anthony at the University of Arkansas. Genomes for AP and AR lines were sequenced and then aligned with reference genome Red Jungle Fowl. Unique SNPs were identified in AP and AR lines to be potential indicators associated with progression or regression of RSV-induced tumors. Over 7 million SNPs were filtered based on frequency, location, read depth, and gene function. A total of 9 candidate SNPs, 4 for AP and 5 for AR, were chosen as candidate biomarkers associated with RSV tumor growth.

Genomic sequencing results were based on 10 pooled DNA samples for each line; therefore SNPs were validated in larger populations of 96 per line. SNPs showing high frequency differences

between AP and AR lines were compared, and those with highest divergence were considered for further validation. Genotyping results showed SNPs SNPs between AP and AR lines were at least 60% or more divergent. The average SNP frequency for all 9 candidates was 92%, meaning the AP and AR lines were 92% different when considering all 9 SNP locations.

To investigate any genotype-phenotype associations, two un-related populations were challenged with RSV and genotyped at the 9 SNP locations. The two un-related populations used were the Giant jungle fowl (GJF) and White Leghorn (WL). Birds within each line were grouped based on phenotype RES or SUS, where RES birds are those whose tumors regressed completely and SUS birds were those whose tumors progressed drastically. The RES and SUS groups for GJF and WL lines were genotyped. There was no significant difference found between RES and SUS birds in either chicken line. SNP frequencies between RES and SUS birds were almost identical, showing no genotypic correlation to the phenotypic trait of interest. Because the AP and AR lines have been intensely selected based on one phenotype, genetic drift caused from bottlenecking could have segregated the alleles at the 9 SNP locations studied. Further testing with a larger number of candidate SNPs in larger populations could reveal sets of haplotypes associated with tumor destiny.

Although no SNP differences were detected between RES and SUS birds within each line, there was a significant difference ($p=0.0001$) detected for grand mean tumor score between the GJF and WL lines. The RJF is theorized to be the ancestral bird of all chicken breeds. The SNP differences between GJF and WL lines could be evidence of domestication as GJF is more closely related to RJF than WL. GJF showed a grand mean 33% difference to RJF based on the 9 SNPs and WL showed a grand mean 60% difference to RJF based on the 9 SNPs. To further tests these SNPs for evidence as chicken domestication, larger samples sizes and additional chicken breeds would need to be used.

The AR and AP lines have been heavily selected for 60 years or more, which would explain the complete genetic divergence between the two lines. The GJF population has most likely experienced strong random genetic drift due to random selection while maintained at the AES. With over 12,000 unique SNPs found between the AR and AP lines, there are a large number of haplotype combinations

that could be responsible for the response to Rous sarcoma virus in chickens. Only nine out of that 12,000 were chosen as candidate biomarkers based on gene function and SNP frequency. There are many more opportunities to continue studying the genotype-phenotype relationship associated with RSV-induced tumors.

UNIVERSITY OF ARKANSAS

Research Compliance
Office of the Director

MEMORANDUM

TO: Nick Anthony

FROM: Craig N. Coon, Chairman
Institutional Animal Care
And Use Committee

DATE: February 14, 2011

SUBJECT: IACUC PROTOCOL **APPROVAL**
Expiration date : **February 6, 2014**

The Institutional Animal Care and Use Committee (IACUC) has **APPROVED** Protocol **#11025- "EVALUATION OF THE DEVELOPMENT AND REGRESSING/PROGRESSION OF ROUS SARCOMA TUMORS IN ARKANSAS REGRESSOR AND PROGRESSOR CHICKENS."** You may begin this study immediately.

In granting its approval, the IACUC has approved only the protocol provided. Should there be any changes in the protocol during the research, please notify the IACUC in writing **prior** to initiating the changes. If the study period is expected to extend beyond **02-6-2014,** you must submit a new protocol. By policy the IACUC cannot approve a study for more than 3 years at a time.

The IACUC appreciates your cooperation in complying with University and Federal guidelines for research involving animal subjects.

cnc/car

cc: Animal Welfare Veterinarian

120 Ozark Hall • 1 University of Arkansas • Fayetteville. AR 72701
Voice (479) 575-3845 • Fax (479) 575-3846